

2015

# Considerations for Screening Designs and Follow-Up Experimentation

Robert D. Leonard

*Virginia Commonwealth University*, [leonardrd@vcu.edu](mailto:leonardrd@vcu.edu)

Follow this and additional works at: <http://scholarscompass.vcu.edu/etd>

 Part of the [Statistics and Probability Commons](#)

© The Author

---

Downloaded from

<http://scholarscompass.vcu.edu/etd/3928>

This Dissertation is brought to you for free and open access by the Graduate School at VCU Scholars Compass. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of VCU Scholars Compass. For more information, please contact [libcompass@vcu.edu](mailto:libcompass@vcu.edu).

CONSIDERATIONS FOR SCREENING DESIGNS AND FOLLOW-UP EXPERIMENTATION

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

by

Robert D. Leonard  
B.S.Ed., Ohio University, 2003

Director: David J. Edwards, Associate Professor, Statistical Sciences and Operations  
Research

Virginia Commonwealth University  
Richmond, Virginia  
August, 2015

## Acknowledgment

I would like to thank my advisor, Dr. Edwards, for his time and encouragement. I would also like to thank my committee members for providing a positive influence on the work presented in this dissertation and to thank my friends and colleagues in the SSOR department their support. I would also like to thank the National Institute of Aerospace and the NASA Langley Research Center for the funding support and the opportunity to work on challenging problems. Lastly but most importantly, I would like to thank my wife, Gisel. Because of your love and understanding, we have been able to take the road less travelled and enjoy what there is to learn along the way.

# Table of Contents

List of Tables .....	vi
List of Figures .....	vii
Abstract.....	ix
<b>Chapter</b>	
<b>1 Introduction</b>	
1.1 Motivation .....	1
1.2 Dissertation Topics	
1.2.1 Optimum Screening Designs with & without Partial Replication .....	4
1.2.2 Screening Design and Protecting Against Model Misspecification .....	4
1.2.3 Considerations for Follow-Up Experimentation .....	5
1.3 Research Objectives .....	5
1.4 Dissertation Outline .....	6
<b>2 Background and Literature Review</b>	
2.1 Introduction .....	7
2.2 Model Setting .....	7
2.3 Model Selection Procedures	
2.3.1 A Modified Forward Selection Procedure .....	12
2.3.2 The Dantzig Selector .....	15
2.4 Optimality Criteria .....	19
2.5 Simulation Setting .....	23
2.6 Design Algorithms	
2.6.1 Candidate Exchange Algorithm .....	24
2.6.2 Coordinate-Exchange Algorithm .....	26
2.6.3 The Pareto Aggregating Point Exchange Method .....	28

<b>3 Optimum Screening Designs with and without Partial Replication</b>	
3.1 Introduction .....	30
3.2 <i>DP</i> -Optimality .....	31
3.3 A Bayesian Framework	
3.3.1 Bayesian <i>DP</i> -Optimality .....	32
3.3.2 Constructing Designs with a Fixed Number of Replicate Points .....	36
3.3.3 Comparing Properties of Designs with Varying $df_{pe}$ .....	37
3.4 Simulation Results .....	43
3.4.1 Results for 7-factor designs with $n = 20$ runs .....	44
3.4.2 Results for 7-factor designs with $n = 32$ runs.....	51
3.4.3 Results for 10-factor designs with $n = 20$ runs.....	57
3.5 Final Comments and Discussion .....	63
<b>4 Screening Methods &amp; Protecting Against Model Misspecification</b>	
4.1 Introduction .....	65
4.2 Model-Robust Factorial Designs .....	66
4.3 The Pareto Front Approach	
4.3.1 The Pareto Front Approach - An Example from the Literature .....	67
4.3.2 Considering the Amount of Replication when Choosing a Design .....	72
4.3.3 PAPE BD Screening Designs .....	74
4.3.4 PAPE BDP Screening Designs .....	76
4.3.4 Comparing BDP and PAPE BDP Designs .....	79
4.4 Simulation Results .....	80
4.5 Comments and Discussion .....	86
<b>5 Considerations for Follow-Up Experimentation</b>	
5.1 Introduction .....	88
5.2 Optimal Determination of Follow-Up Runs	
5.2.1 <i>D</i> -Optimal Follow-Up .....	89
5.2.2 Bayesian <i>DP</i> -Optimal Follow-Up .....	90
5.3 Simulation Settings for Two-Stage Sequential Experimentation .....	93

5.4 Simulation Results	
5.4.1 8-factor Designs using $n_i = 16$	98
5.4.2 7-factor Designs using $n_i = 16$	90
5.5 Comments and Discussion	108
<b>6 Conclusion and Final Comments</b>	<b>110</b>
Reference List	114
Appendix	118

## List of Tables

3.1	Formulations for criterion efficiencies .....	38
3.2	Characteristics of 7-factor 20-run designs .....	43
3.3	Table 3.3 Characteristics of 7-factor 32-run designs .....	52
3.4	Characteristics of 10-factor 20-run designs .....	58
4.1	$D_{\text{eff}}$ and $\text{tr}(\mathbf{A}\mathbf{A}')$ values for Pareto front example .....	69
4.2	Estimation capacities for 8-factor PAPE BD, PAPE BDP, and MRFD's .....	78
5.1	Un-replicated and partially replicated 8-factor MRFD's .....	95
5.2	Initial and follow-up designs for eight factors .....	96
5.3	Un-replicated and partially replicated 8-factor MRFD's .....	105

## List of Figures

2.1	Correlation structures of regular and non-regular designs .....	11
2.2	Example profile plot when implementing the Dantzig selector .....	18
2.3	Candidate-exchange algorithm example .....	25
2.4	Coordinate-exchange algorithm example .....	26
2.5	Three PAPE search methods (Lu & Anderson-Cook, 2013).....	29
3.1	Greyscale correlation maps for 7-factor, 20-run non-replicated designs .....	35
3.2	Efficiencies for 7-factor, 20-run designs.....	39
3.3	Efficiencies for 7-factor, 32-run designs.....	40
3.4	Efficiencies for 10-factor, 20-run designs.....	40
3.5	Correlation maps for BDP-optimal Design and its unique runs .....	41
3.6	Scatterplots of $\text{tr}(\mathbf{AA}')$ versus number of replicates for 7-factor designs .....	42
3.7	Scatterplots of $\text{tr}(\mathbf{AA}')$ versus number of replicates for 10-factor designs .....	43
3.8	Figure 3.8 Cover and FDR's when 2/7 main effects are active ( $n=20$ ) .....	48
3.9	Cover and FDR's when 4/7 main effects are active ( $n=20$ ) .....	49
3.10	Number of terms in final model ( $k=7, n=20$ ) .....	49
3.11	Power and type I error rates for effects ( $k=7, n=20$ ).....	50
3.12	FDR's for main effects and two-factor interactions ( $k=7, n=20$ ).....	50
3.13	Cover and FDR's when 2/7 main effects are active ( $n=32$ ) .....	55
3.14	Cover and FDR's when 4/7 main effects are active ( $n=32$ ).....	55
3.15	Number of terms in final model ( $k=7, n=32$ ) .....	56
3.16	Power and type I error rates for effects ( $k=7, n=32$ ) .....	56
3.17	FDR's for main effects and two-factor interactions ( $k=7, n=32$ ).....	57
3.18	Correlation map of 10-factor Bayesian D-optimal design ( $n=20$ ) .....	58
3.19	Cover and FDR's when 2/10 main effects are active ( $n=20$ ) .....	61



3.20	Cover and FDR's when 8/10 main effects are active (n=20) .....	61
3.21	Number of terms in final model (k=10, n=20).....	62
3.22	Power and type I error rates for effects (k=10, n=20) .....	62
3.23	FDR's for main effects and two-factor interactions (k=10, n=20) .....	63
4.1	Illustration of Pareto front in two dimensions .....	68
4.2	Plot indicating distribution of weightings for example designs.....	70
4.3	Pareto front results depicted in Lu et al. (2011) .....	71
4.4	Reproduced results from Lu et al. (2011) colored by $df_{pe}$ .....	73
4.5	Scatterplot matrices of criterion values for PAPE designs .....	74
4.6	Mixture plots for PAPE BD designs .....	76
4.7	Mixture plots for PAPE BDP designs .....	77
4.8	Estimation capacities for PAPE BDP designs .....	78
4.9	Boxplots of $tr(\mathbf{AA}')$ for BDP and PAPE BDP Designs .....	79
4.10	Correlation maps of BD designs with $tr(\mathbf{AA}')$ minimized .....	81
4.11	Cover and FDR's when 2/8 main effects are active (n=16) .....	83
4.12	Cover and FDR's when 4/8 main effects are active (n=16) .....	84
4.13	Number of terms in final model (k=8, n=16) .....	84
4.14	Power and type I error rates for effects (k=8, n=16) .....	85
4.15	FDR's for main effects and two-factor interactions (k=8, n=16) .....	85
5.1	Correlation maps for 8-factor initial designs .....	95
5.2	Correlation maps of initial design and complete design .....	98
5.3	Average number of $df_{pe}$ and RMSE for 8-factor case .....	99
5.4	Power and type I error rates for 8-factor case .....	100
5.5	Coverages by effect size for 8-factor case .....	101
5.6	Number terms fit, cover, and FDR's for 8-factor case .....	102
5.7	Correlation maps for 7-factor initial designs .....	104
5.8	Number terms fit, cover, and FDR's for 7-factor case .....	106
5.9	Power, type I error, FDR's for 7-factor case .....	107
5.10	Cover, power, and FDR's for moderate effect sizes for 7-factor case .....	107

## Abstract

### CONSIDERATIONS FOR SCREENING DESIGNS AND FOLLOW-UP EXPERIMENTATION

By Robert D. Leonard

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

Virginia Commonwealth University, 2015.

Major Director: David J. Edwards, Associate Professor, Statistical Sciences and Operations Research

The success of screening experiments hinges on the effect sparsity assumption, which states that only a few of the factorial effects of interest actually have an impact on the system being investigated. The development of a screening methodology to harness this assumption requires careful consideration of the strengths and weaknesses of a proposed experimental design in addition to the ability of an analysis procedure to properly detect the major influences on the response. However, for the most part, screening designs and their complementing analysis procedures have been proposed separately in the literature without clear consideration of their ability to perform as a single screening methodology.

As a contribution to this growing area of research, this dissertation investigates the pairing of non-replicated and partially-replicated two-level screening designs with model selection procedures that allow for the incorporation of a model-independent error estimate. Using simulation, we focus attention on the ability to screen out active effects from a first

order with two-factor interactions model and the possible benefits of using partial replication as part of an overall screening methodology. We begin with a focus on single-criterion optimum designs and propose a new criterion to create partially replicated screening designs. We then extend the newly proposed criterion into a multi-criterion framework where estimation of the assumed model in addition to protection against model misspecification are considered. This is an important extension of the work since initial knowledge of the system under investigation is considered to be poor in the cases presented. A methodology to reduce a set of competing design choices is also investigated using visual inspection of plots meant to represent uncertainty in design criterion preferences. Because screening methods typically involve sequential experimentation, we present a final investigation into the screening process by presenting simulation results which incorporate a single follow-up phase of experimentation. In this concluding work we extend the newly proposed criterion to create optimal partially replicated follow-up designs. Methodologies are compared which use different methods of incorporating knowledge gathered from the initial screening phase into the follow-up phase of experimentation.

# Chapter 1

## Introduction

### 1.1 Motivation

Controlled, repeatable results are the foundation to scientific inquiry even in cases where limited resources do not permit a full investigation of all conceivable scenarios believed to influence a system. If possible, experimentation is usually carried out as a sequential process (Goos and Jones, 2011) where, over time, a cycle of events allows researchers and subject matter experts to refine and re-design experiments as their understanding of a problem matures. The initial screening phase of this sequence typically involves using a limited number of experimental runs to screen out the most influential effects. Unfortunately, this screening process does not always lead to a definitive model due to aliasing, or confounding, of effects. In such situations, additional follow-up runs can be added to the initial experiment in order to make more informed decisions concerning which effects are truly having the most influence.

In addition to powerful experimental designs, analysis procedures are also required that can properly analyze the results (Abraham (1999), Edwards and Mee (2011), Li and Lin (2003)). Due to the limited number of observations typically collected in screening experiments, complete or partial aliasing of effects can lead many model selection procedures, and consequentially their users, astray. Therefore, improving our understanding of the combined ability of screening designs and model selection methods is needed for such approaches to improve, become more reliable, and gain further acceptance

in practice (Anderson-Cook and Hamada (2014), Scinto et al. (2014)). However, determining how to identify and control for the impact of these aliasing structures is not trivial, especially when considering the possible limitations of different model selection methods and when considering the wide array of applications of experimental design.

To identify the potential of a complete screening methodology's ability to properly detect truly active effects requires simulation (Brenneman (2014)), which can be computationally intensive. Further, a number of designs and analysis procedures have been proposed for screening purposes yet the ability to simulate their combined potential via more efficient computer routines has become a more recent possibility, hence research in this area is only beginning to take form (Draguljic et al. (2014), Marley and Woods (2010), Mee et al. (2015)). As an extension of this recent research, this dissertation incorporates partial replication into initial screening and follow-up designs and investigates the ability of analysis procedures to benefit from a model-independent error estimate.

Bayesian methods are employed in order to allow the estimation of potentially active effects not included in the assumed model; however, the methods employed are more of a classical approach used to form designs less dependent on assumed inactive effects. We use this approach to create partially replicated initial screening designs via a newly proposed optimality criterion in addition to proposing a two-stage approach for creating designs with a fixed number of replicate runs. Possible advantages to using partial replication are presented along with recommendations based upon simulation study results using two-level designs in conjunction with two model selection procedures – the Dantzig selector (Phoa et al. (2009) and a modified forward selection approach (Mee et al. (2015)). These two methods are used due to their popularity in practice and ease of use.

We focus on the optimal design of experiments since the approach provides a flexible alternative to more traditional approaches when creating and augmenting screening designs (Atkinson et al. (2007), Goos and Jones (2011), Silvestrini (2013)). To fully utilize such optimization procedures requires careful consideration of *a priori* knowledge of the system before starting the investigation. However, in screening situations, we often lack any knowledge concerning the influence of any particular effect. Therefore, multiple design criteria should be considered when creating or selecting an experimental design to permit to information concerning the system model to guide the choice of design points while still maintaining some protection against model misspecification (Edwards and Kelly (2014), Lu et al., (2011)). While we initially focus on a single-criterion approach to create screening designs in Chapter 3, Chapter 4 allows for a consideration of multiple criteria.

As a final comparison of selected designs being paired with model selection procedures, we employ the use of simulation in each chapter of this dissertation. To compare the performances of screening methodologies to detect the most influential effects, our simulation results present metrics in terms of effect size (dominant, moderate, and small) in addition to effect type (main effects or two-factor interactions). This allows for a comparison in ability of methods to specifically detect the most influential active effects even if some two-factor interactions are having more of an influence than some main effects. Using a combination of these metrics allows for a more informative analysis of simulation results. We also present results illustrating the number of fitted model terms since they help to highlight some of the weaknesses and strengths of certain combinations of design and analysis procedures.

## 1.2 Dissertation Topics

### 1.2.1 Optimum Screening Designs with & without Partial Replication

We propose a Bayesian modification of a design criterion used to create partially replicated experimental designs. We also provide details concerning a two-step coordinate/candidate-exchange algorithm used to construct partially replicated screening designs. This two-stage procedure permits direct control over the number of desired design points to be replicated. Using simulation, we compare performances of partially replicated and non-replicated designs when using the proposed modification.

### 1.2.2 Screening Design and Protecting Against Model Misspecification

Although  $D$ -optimality is the single most popular optimality design criterion used in practice, a number of  $D$ -optimal designs can result from an algorithmic search. However, further ranking of designs can take place by considering additional criteria. Further, in some situations, such as in screening where *a priori* knowledge of the model is poor, a less than  $D$ -optimal choice of design may actually be better suited to the application. We introduce a multi-criterion approach for creating a suite of competing designs using combinations of criteria meant to protect against potential harm due to model misspecification. We then examine methods involving the visual inspection of plots used to reduce a set of competing design choices. Simulation study results are then used to compare performances of a final choice of designs when using these plots.

### **1.2.3 Considerations for Follow-Up Experimentation**

We extend the comparison of non-replicated and partially replicated screening methodologies to a sequential experimentation framework. We begin by providing additional background information concerning optimal approaches to augmenting initial designs with follow-up runs. We then propose an approach which allows for the optimal determination of partially replicated follow-up runs. Simulation results are then used to show the possible benefits of following-up initially un-replicated designs with a set of partially replicated runs.

## **1.3 Research Objectives**

The common objective of this dissertation is to identify and demonstrate best practices that provide practical improvements to the overall screening process. This includes an investigation of experimental design, analysis procedures and the consideration of sequential experimentation. Simulation results are presented in Chapters 3 and 4 comparing the performance of forward selection and the Dantzig selector when paired with competing design choices. The benefit of a model-independent error is investigated by considering un-replicated and partially replicated screening designs. In Chapter 5, the use of partially replicated follow-up designs are investigated when only implementing forward selection.

In Chapter 3 we will present alternative approaches to creating partially replicated optimum screening designs to those found in the literature when a single criterion is used and provide supporting evidence for their consideration. In Chapter 4 we will show that the flexibility of partially replicated optimum designs can be extended to considerations for model-misspecification by examining a reduced set of designs using a multi-criterion



approach. We will also demonstrate that care should be taken when using a particular design criterion since an unintended consequence can be the creation of designs with a large number of replicate points. In Chapter 5 we introduce an optimal approach to augmenting initial screening designs that allows for the determination of replicate follow-up design points. We will show that even when initial understanding of the true underlying model is poor, information from the initial experiment can be used to help identify active effects.

## **1.4 Dissertation Outline**

Chapter 2 provides background information and literature review covering the fundamental topics used throughout this dissertation. Chapters 3, 4, and 5 each provide additional background information for the topic covered, present simulation results from implementing the proposed methodology, and conclude with a discussion of the overall results. Chapter 6 provides final concluding comments and discussion of planned future work.

# Chapter 2

## Background and Literature Review

### 2.1 Introduction

In this chapter we set notation and provide a review of the fundamental topics used in this dissertation. First, the model setting and its connection to experimental design are presented. Next, details concerning a modified forward selection procedure and the Dantzig selector are provided. Then, we provide a summary of optimality criteria used in creating and evaluating screening designs. We then provide details concerning the simulation setting and end with a summary of algorithms used in this dissertation to create designs.

### 2.2 Model Setting

We consider the linear model written in matrix form as

$$\mathbf{y} = \beta_0 \mathbf{1} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (2.1)$$

where  $\mathbf{y}$  is an  $n \times 1$  response vector of observations,  $\beta_0$  is the intercept term,  $\mathbf{1}$  is an  $n \times 1$  vector of ones,  $\mathbf{X}$  is an  $n \times p$  model matrix,  $\boldsymbol{\beta}$  is  $p \times 1$  vector of unknown parameters, and assume  $\boldsymbol{\epsilon} \sim N(0, \sigma^2 \mathbf{I}_n)$  is an  $n \times 1$  vector of random errors. We also denote the number of experimental runs, or sample size, as  $n$ . In this dissertation we focus on main-effects and two-factor interactions influencing the response. That is, (2.1) takes the equation form

$$\mathbf{y} = \beta_0 + \sum_{i=1}^k \beta_i X_i + \sum_{i=1}^{k-1} \sum_{j=i+1}^k \beta_{ij} X_i X_j + \boldsymbol{\epsilon} \quad (2.2)$$

where  $k$  is the number of factors, or main-effects, being investigated. The ability to properly estimate this type of model is problematic when resources are limited due to the number of two-factor interactions growing at a rate of  $\binom{k}{2} = k(k-1)/2$  as the  $k$  number of factors being considered increases. That is, when resources limit the number of experimental runs such that  $n < p$ , we cannot simultaneously estimate all  $p = k + 1 + \binom{k}{2}$  terms in model (2.2) since effects are completely or at least partially aliased with each other. For illustrative purposes only, consider the following simple example. Suppose  $k = 3$  factors are to be investigated, but budgetary constraints will not allow for more than  $n = 4$  experimental runs to be carried out. Therefore the full factorial  $2^3$  design using all possible 2-level combinations in 8 experimental runs cannot be used. However, say we plan to fit the main-effects only model, which for this example would be

$$y = \beta_0 + \sum_{i=1}^3 \beta_i X_i + \epsilon.$$

If we can safely ignore two-factor interactions as possibly being active, one possible experimental design in four runs which allows for estimation of all three factors and the intercept without having aliasing problems could be

$$\mathbf{D} = \begin{bmatrix} -1 & -1 & 1 \\ 1 & -1 & -1 \\ -1 & 1 & -1 \\ 1 & 1 & 1 \end{bmatrix}. \quad (2.3)$$

This simple design is created by using a one-half fraction of the  $2^3$  design. That is, only half of the  $2^3 = 8$  experimental runs of all possible 2-level combinations are used to create the  $2^{3-1}$  fractional factorial design (FFD) shown in (2.3). Further, a design generator  $I = x_1 x_2 x_3$ , where  $I$  is the identity column, is specifically chosen to create the columns in order to permit

all main effects to be un-aliased and therefore independently estimable. We can visually inspect the aliasing structure using the information matrix,  $\mathbf{X}'_1\mathbf{X}_1$ , corresponding to the model matrix  $\mathbf{X}_1 = [\mathbf{1}|\mathbf{D}]$  and see that it is orthogonal. That is,

$$\mathbf{X}'_1\mathbf{X}_1 = \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix},$$

where all the off-diagonal elements are zeros; however, if two-factor interactions could not be safely ignored, then aliasing of particular main effects with particular two-factor interactions occurs. Using the generator above, we can see the exact relationship between the fully aliased effects, where

$$x_1 = x_2x_3, \quad x_2 = x_1x_3, \quad x_3 = x_1x_2. \quad (2.4)$$

This can be easily visualized for this simple example by also inspecting the information matrix when combining the original main-effects only model matrix,  $\mathbf{X}_1$ , with the matrix  $\mathbf{X}_2$ , which contains the three additional two-factor interactions ( $x_1x_2$ ,  $x_1x_3$ , and  $x_2x_3$ ) left out of the model. That is,

$$\begin{aligned} \mathbf{X}'\mathbf{X} &= \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} [\mathbf{X}_1|\mathbf{X}_2] \\ &= \begin{bmatrix} 4 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 & 0 & 0 & 4 \\ 0 & 0 & 4 & 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 4 & 4 & 0 & 0 \\ 0 & 0 & 0 & 4 & 4 & 0 & 0 \\ 0 & 0 & 4 & 0 & 0 & 4 & 0 \\ 0 & 4 & 0 & 0 & 0 & 0 & 4 \end{bmatrix}. \end{aligned}$$

Here we can see that the off diagonals are no longer zeros. In fact, because the off diagonals are equal to  $n = 4$  we know that there is full aliasing between some effects. By examining the upper right portion of  $\mathbf{X}'\mathbf{X}$ , namely

$$\begin{aligned}
\mathbf{X}'_1\mathbf{X}_2 &= \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 & -1 \\ -1 & -1 & 1 \\ -1 & 1 & -1 \\ 1 & 1 & 1 \end{bmatrix} \\
&= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 4 \\ 0 & 4 & 0 \\ 4 & 0 & 0 \end{bmatrix},
\end{aligned}$$

the full aliasing given in (2.4) using the design generator can be seen when comparing corresponding rows and columns.

When designs create this type of aliasing structure (i.e., effects are either orthogonal or completely confounded) they are known as *regular* designs. In this dissertation we focus on *non-regular* designs, which create partial aliasing of effects. Although partial aliasing can be problematic due to the complex aliasing structures created, we employ the use of these designs since there is more of a possibility that information concerning the aliased effects can be used for estimation purposes and they also allow for more flexibility in creating designs of different run sizes (Montgomery (2009), Wu and Hamada (2009)). To illustrate the benefit and drawback to this approach, let us continue with the example and assume that the budget allowed for one additional experimental run. Since  $2^{k-p}$  designs can only be created for  $1/2^p$  fractions of  $2^k$  designs, a design of size  $n = 5$  runs cannot be created using this approach. However, implementing optimal design of experiments allows designs of any run size to be created. Take the design shown in (2.5) for example, created using the Bayesian  $D$ -optimality criterion, which will be explained in the next chapter.

$$\mathbf{D}_{opt} = \begin{bmatrix} -1 & -1 & 1 \\ 1 & -1 & -1 \\ -1 & 1 & -1 \\ 1 & 1 & 1 \\ -1 & 1 & 1 \end{bmatrix} \tag{2.5}$$

Comparing the 4-run FFD from (2.3) to the 5-run non-regular design in (2.5), it can be seen that  $\mathbf{D}_{opt}$  is the FFD given in (2.3) with the addition of a fifth experimental run  $[-1 \ 1 \ 1]$ . In fact, using the same criterion to form a four-run design creates the FFD in (2.3), illustrating the flexibility in using optimal design of experiments. However, it should be noted that this is a simplistic example and implementing optimal design of experiments does not always result in producing regular designs for any given run size (DuMouchel and Jones, 1994). Further, let us return to the complex aliasing that can result when using non-regular designs. Greyscale maps and effect correlation values for the two example designs are shown in Figure 2.1. The regular design structure of the FFD is easily apparent on the left. The 5-run non-regular design on the right shows the partial aliasing that can result, where every effect is at least partially aliased with another effect or sets of effects and the main effects are no longer orthogonal to each other.

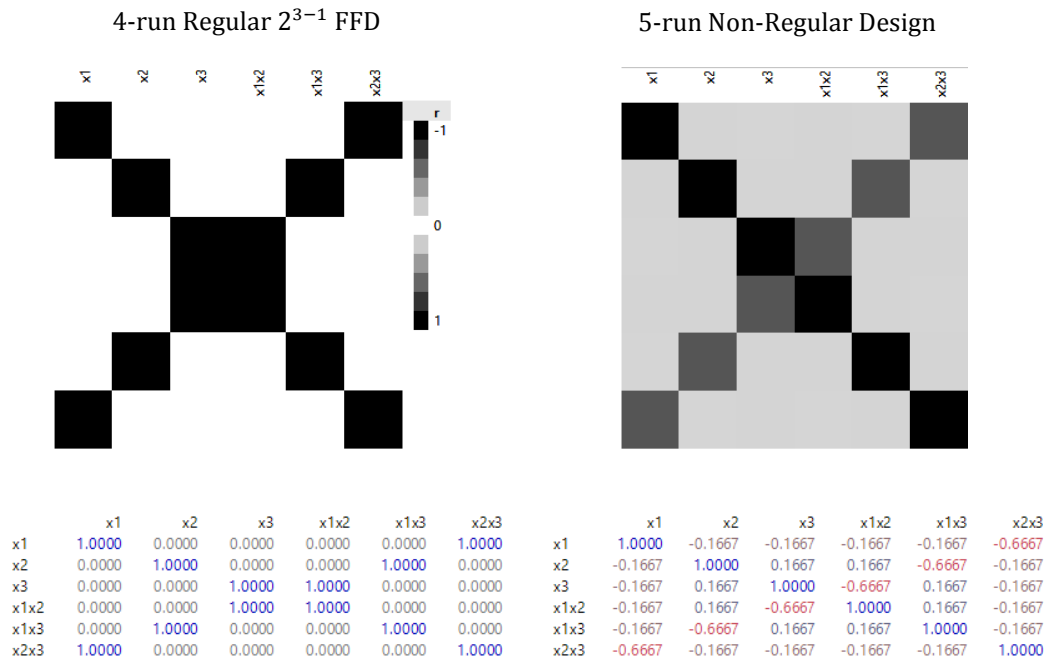


Figure 2.1 Correlation structures of regular and non-regular designs.

The messy aliasing structure of the non-regular design produces the following relationships

$$\begin{aligned}x_2 &= x_3 - x_1x_2 + x_1x_3 \\x_1 &= -x_3 + x_1x_2 + x_2x_3\end{aligned}\tag{2.6}$$

Although in (2.6) there is no full aliasing of effects as in (2.2) for the FFD, the structure of the non-regular design can still lead to problems with analysis procedures. For example, suppose the true model were  $y = -8x_2 + 7x_3 + 5x_1x_3$ . Because of the top relation in (2.6), a combination of all true model effects is aliased with a single inactive  $x_1x_2$  interaction effect, which can cause analysis procedures to incorrectly identify the  $x_1x_2$  effect as active. Rules and model building principles can be incorporated into analysis procedures to help reduce the negative impact from aliasing, some of which are incorporated into the procedures used in this dissertation; however, constructing a design that complements an analysis procedure should also be a consideration. Therefore, for screening situations, we seek to construct designs which minimize the impact of these aliasing structures so that analysis of the experimental results using model selection procedures can more efficiently detect the active effects.

## 2.3 Model Selection Procedures

### 2.3.1 A Modified Forward Selection Procedure

Forward selection (FS) procedures have long been known for their inability to control for over-fitting a model, especially in cases where there are many competing models (Abraham et al. (1999), Edwards and Mee (2011), Lin (1993) and comment by Wang (1995)). However, the FS procedure is widely available in software packages and easy to implement. Further,

modifications such as controlling experiment-wise error rate (EER) via Bonferroni adjusted  $p$ -values to help control over-fitting (Westfall et al., 1997) and the incorporation of a model-independent error estimate to improve performance (Mee at al., 2015) have also strengthened the procedure. To illustrate the calculation of a model-independent, or pure error estimate, consider the following partitioning of total variability into regression and residual sums of squares by

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2.7)$$

where  $\hat{y}_i$  is the  $i$ th predicted response value from  $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$  using the usual least squares estimates

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}. \quad (2.8)$$

Beginning with an empty model, FS sequentially adds one candidate term at a time until converging to a solution model. The FS procedure relies upon  $Var(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ , the variance-covariance matrix for the usual least squares estimates, to determine which effect should be added at each step. Partial  $t$ -tests and corresponding  $p$ -values are formed using the test statistic

$$t_{obs} = \frac{\hat{\beta}_j}{\hat{\sigma} \sqrt{c_{jj}}} \sim t_{n-p} \quad (2.9)$$

where  $c_{jj}$  is the  $j$ th diagonal element of  $(\mathbf{X}'\mathbf{X})^{-1}$ . Therefore, the addition of terms to a model using FS is a component of an estimate of  $\sigma^2$  as well as a component of the design,  $\mathbf{X}$ . That is, a design should be used which minimizes  $c_{jj}$ , and in cases where a model-dependent error is used, a design should be used which also provides the best ability to estimate  $\sigma^2$ .



A model-dependent estimate of  $\sigma^2$  can be formed by dividing the residual sums of squares in (2.7) by its degrees of freedom, giving mean square error (MSE)

$$\hat{\sigma}^2 = MSE = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - p} \quad (2.10)$$

as an estimate of  $\sigma^2$ . Alternatively, a model-independent estimate of  $\sigma^2$  can be formed by considering a further partition of the residual sums of squares in (2.7). This can be done by considering the grouping of repeated response observations (Myers, pg. 117). Let there be  $c$  combinations of regressor variables with  $n_i$  experimental runs at the  $i$ th combination. Let  $y_{ij}$  be the  $j$ th response at the  $i$ th combination where  $i = \{1, 2, \dots, c\}$ ,  $j = \{1, 2, \dots, n_i\}$ , and  $\sum_{i=1}^c n_i = n$  total experimental runs. Then, consider the partitioning of the residual sums of squares as

$$\sum_{i=1}^c \sum_{j=1}^{n_i} (y_{ij} - \hat{y}_i)^2 = \sum_{i=1}^c \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 + \sum_{i=1}^c n_i (\bar{y}_i - \hat{y}_i)^2, \quad (2.11)$$

where  $\hat{y}_i$  is the predicted response value for the  $i$ th combination of the regressor variables and  $\bar{y}_i$  is the mean response for all  $n_i$  runs at the  $i$ th combination. The first term on the right hand side of (2.11) measures the variation due to repeated observations and provides a mean square *pure error*

$$\hat{\sigma}_{PE}^2 = MSE_{pure} = \sum_{i=1}^m \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 / \sum_{i=1}^m (n_i - 1) \quad (2.12)$$

as an estimate of  $\sigma^2$ . Both (2.10) and (2.12) are incorporated into the FS procedure used in this dissertation. That is, if paired with a partially replicated design, FS using  $\hat{\sigma}_{PE}^2$  in (2.12) can be used to conduct  $t$ -tests using (2.9), otherwise  $\hat{\sigma}^2$  as given in (2.10) is used when FS is paired with an un-replicated design. The FS procedure used in this dissertation also makes

use of a Bonferroni adjustment (Westfall, 1998) of the  $p$ -values resulting from the  $t$ -tests by (2.9) in order to help control the tendency of FS to overfit models. This is implemented by multiplying each naïve individual  $p$ -value by the number of eligible terms at each step of the procedure to form an experiment-wise error rate (EER). The procedure of adding terms to the model terminates when the adjusted  $p$ -value first exceeds a pre-specified EER. We use  $\text{EER}=0.50$  for all applications of FS, as suggested by Mee (2013).

Another benefit to using the FS procedure is that it provides the user the ability to account for a weak or strong effect heredity assumption (Mee, 2013). That is, in applications where two-factor interactions are considered, only interactions conforming to pre-specified heredity assumptions are allowed to enter the "current" model as the FS procedure adds one variable at a time. A model with *strong* heredity follows the property that if it contains an interaction effect  $x_i x_j$  then it must also contain both of its parent effects  $x_i$  and  $x_j$ . A model with *weak* heredity follows the property that if it contains an interaction effect  $x_i x_j$  then it must also contain at least one of its parent effects. In this dissertation, we consider weak effect heredity for all simulated responses and incorporate this assumption into the FS procedure. Thus, the first term to enter the model will always be a main effect, and only two-factor interactions with at least one parent effect already in the model are given consideration in subsequent steps of the process.

### 2.3.2 The Dantzig Selector

Candes and Tao (2007) proposed the Dantzig selector as an estimator of model parameters in situations where the number of experimental runs  $n$  is less than the number of model parameters of interest  $p$  (i.e., supersaturated cases). Let  $\mathbf{X}$  be a  $n \times p$  model matrix with

columns unit-normed. That is, the columns of  $\mathbf{X}$  have the same Euclidean size. Then, the Dantzig estimator  $\hat{\boldsymbol{\beta}}$  is the solution to the  $l_1$ -regularization problem

$$\begin{aligned} \min_{\hat{\boldsymbol{\beta}} \in \mathbf{R}^k} \|\hat{\boldsymbol{\beta}}\|_{l_1} \\ \text{s. t. } \|\mathbf{X}'\mathbf{r}\|_{l_\infty} \leq \delta, \end{aligned} \quad (2.13)$$

where  $\mathbf{r} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$  is the residual vector,  $\delta$  is the tuning parameter, and for a vector  $\mathbf{a}$ ,  $\|\mathbf{a}\|_{l_1} = \sum \mathbf{a}_i$ , and  $\|\mathbf{a}\|_{l_\infty} = \max |\mathbf{a}_i|$ . Recasting the Dantzig selector as a linear program we then have

$$\begin{aligned} \min \sum_i u_i \\ \text{s. t. } -\mathbf{u} \leq \hat{\boldsymbol{\beta}} \leq \mathbf{u} \end{aligned} \quad (2.14)$$

and

$$-\delta \mathbf{1}_k \leq \mathbf{X}^{*'}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \leq \delta \mathbf{1}_k,$$

where  $\hat{\boldsymbol{\beta}} \in \mathbf{R}^k$  and we aim to optimize the variables  $\mathbf{u}$  and  $\hat{\boldsymbol{\beta}}$ . This is equivalent to the standard linear program

$$\begin{aligned} \min \mathbf{c}'\mathbf{x} \\ \text{s. t. } \mathbf{A}\mathbf{x} \geq \mathbf{b} \text{ and } \mathbf{x} \geq 0 \end{aligned} \quad (2.15)$$

where

$$\mathbf{c} = \begin{pmatrix} \mathbf{1}_k \\ \mathbf{0}_k \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} \mathbf{X}'\mathbf{X} & -\mathbf{X}'\mathbf{X} \\ -\mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{X} \\ 2\mathbf{I}_k & -\mathbf{I}_k \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} -\mathbf{X}'\mathbf{y} - \delta \mathbf{1}_k \\ \mathbf{X}'\mathbf{y} - \delta \mathbf{1}_k \\ \mathbf{0}_k \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} \mathbf{u} \\ \mathbf{u} + \hat{\boldsymbol{\beta}} \end{pmatrix}.$$

Phoa et al. (2009) suggest the following procedure when screening for important effects when implementing the Dantzig selector:

1. Standardize  $\mathbf{y}$  to have mean 0 and the model matrix  $\mathbf{X}$  to have equal column lengths. Compute  $\delta_0 = \max |\mathbf{x}_i' \mathbf{y}|$ , where  $\mathbf{x}_i$  is the  $i$ th column of  $\mathbf{X}$ .
2. Solve the linear program (2.14) or (2.15) to obtain the Dantzig selector parameter estimates  $\hat{\boldsymbol{\beta}}$  for some values of  $\delta$  ranging from 0 to  $\delta_0$ .
3. Make a profile plot of the estimates by plotting  $\hat{\boldsymbol{\beta}}$  against  $\delta$ .
4. Identify important effects by inspection of the profile plot.

For simulation studies, Phoe et al. (2009) also indicate that  $\delta$  can be chosen automatically based upon a model selection criterion such as AIC, mAIC, AICc, or BIC. The Dantzig selector using BIC, or Bayesian Information Criterion (Schwarz, 1978), is used in our simulation studies since Marley and Woods (2010) indicate that it provides better model fitting capabilities for screening purposes than other information criteria. For linear models, BIC is defined as

$$BIC = n \log(RSS/n) + p * \log(n), \quad (2.16)$$

where  $RSS = \sum_{i=1}^n (y_i - \hat{y}_i)^2$  is the residual sum of squares and  $p$  is the number of model parameters. Further, a threshold parameter  $\gamma$  is used as a lower bound for the magnitude of the Dantzig model parameter estimates,  $\hat{\boldsymbol{\beta}}$ , allowable in the model under consideration. Where incorporation of a model independent error estimate for the Dantzig selector is sought, we introduce

$$\delta_{pe} = \hat{\sigma}_{PE} \sqrt{2 \log(p)}, \quad (2.17)$$

where  $\hat{\sigma}_{PE}$  is a root mean square pure error estimate from (2.12).

Figure 2.2 illustrates an implementation of the Dantzig selector with a plot of shrinking model parameter estimates  $\hat{\boldsymbol{\beta}}$  over a range of  $\delta$  when a 7-factor, 20-run design is used to fit a main-effect with two-factor interactions model. Each curve represents shrinkage

of one of the 28 model parameter estimates when 7 main effects and 21 two-factor interactions are considered. The five labeled effects are those deemed active when choosing a 'best' model by minimum BIC, indicated by  $\delta^*$  in the figure. The literature highly recommends visual inspection of such a plot by subject matter experts in order to ascertain the best model. It is shown in Figure 2.2 that both main effects,  $x_4$  &  $x_7$ , appear to be the most dominate since they tend toward zero slower than any of the other effects. Note also in this example that weak effect heredity appears to be followed. That is, if any of the interaction effects identified as active ( $x_2x_4$ ,  $x_4x_7$ , or  $x_1x_7$ ,) were to be considered in the model, each would have at least one parent effect already in the model since both of the more dominant main effects,  $x_4$  and  $x_7$ , would also likely be included; however, the Dantzig selector does not allow for its direct enforcement and would therefore require direct visual inspection to account for this type of desired model structure.

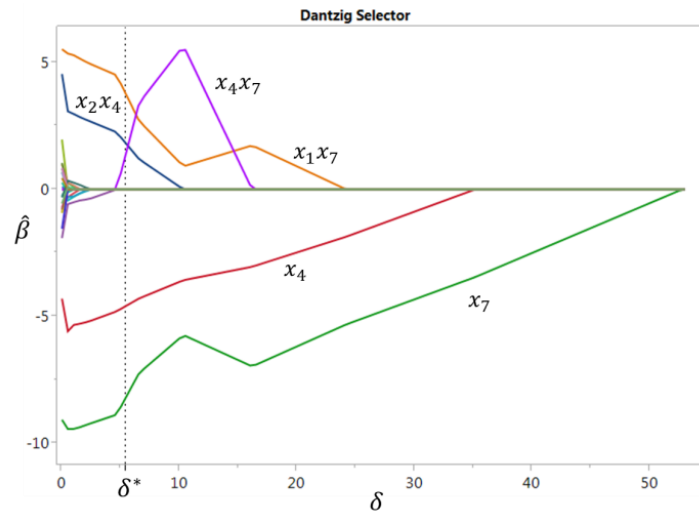


Figure 2.2 Example profile plot when implementing the Dantzig selector

## 2.4 Optimality Criteria

When screening a large number of factors, investigators typically state a main effects only model as their *a priori* model, meaning that they initially assume higher order effects such as two-factor interactions negligible in order to acquire estimates of main effects. Designs created for screening situations where  $n < k$  are known as supersaturated designs. However, in addition to having the ability to estimate main-effects well, we may also seek the ability to estimate any potentially active two-factor interactions. As previously shown,  $n > k$  can easily become  $n < p$  when extending consideration to the estimation of potentially active two-factor interactions. These types of cases, which are known as supersaturated model cases, are the focus of this dissertation. In this section we provide a summary of optimality criteria found in the literature in creating screening designs.

Many theories have been proposed concerning how to create screening designs based upon different design criteria, or functions of measurable design characteristics. One of the first criteria used for systematically creating supersaturated designs, the minimum- $E(s^2)$  criterion, was introduced by Booth and Cox (1962), which minimizes

$$E(s^2) = \left( \sum_{i=1}^{k-1} \sum_{j=i+1}^k s_{ij}^2 \right) / \binom{m}{2} \quad (2.18)$$

where  $s_{ij} = \mathbf{x}_i' \mathbf{x}_j$  is the  $(i,j)$ -th off-diagonal element of the information matrix  $\mathbf{X}'\mathbf{X}$ . The rationale of this criterion is that it can provide a means to create designs with lower correlations between model coefficient estimates and can also serve as a measure of orthogonality.

One of the most widely used criterion, the  $D$ -optimality criterion, is defined as maximizing the determinant of the information matrix,  $\mathbf{X}'\mathbf{X}$ . The history of the criterion can

be traced back to Smith (1918) and the formulation can be attributed to using a determinant to calculate volumes of high dimensional ellipsoids and to  $Var(\hat{\beta}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$  being the variance-covariance matrix for the usual least squares estimates. Wald (1943) used this connection to show that maximizing  $|\mathbf{X}'\mathbf{X}|$  leads to a minimization of the joint confidence ellipsoid, or our uncertainty, concerning  $\hat{\beta}$  and therefore provides a generalized variance estimate for our estimated model coefficients. Later, with the advent of more computing power, algorithms were created to construct  $D$ -optimal designs (Mitchell, 1974), where a  $D$ -optimal design,  $\xi^*$ , can be stated as satisfying

$$|\mathbf{X}(\xi^*)'\mathbf{X}(\xi^*)| = \max_{\xi \in \mathcal{X}} |\mathbf{X}(\xi)'\mathbf{X}(\xi)| \quad (2.19)$$

for all possible designs,  $\xi$ , spanning the entire design space  $\mathcal{X}$ . It is possible for multiple designs to satisfy (2.19), in which case additional criteria may be used to further discriminate between choices of  $D$ -optimal designs. In some situations, it is better to compare designs with respect to  $D$ -optimality by computing their  $D$ -efficiency as

$$D_{eff} = \left( \frac{|\mathbf{X}(\xi)'\mathbf{X}(\xi)|}{|\mathbf{X}(\xi^*)'\mathbf{X}(\xi^*)|} \right)^{\frac{1}{p}} \quad (2.20)$$

where  $p$  is the number of model parameters. The  $A$ -optimality criterion, which minimizes  $tr(\mathbf{X}'\mathbf{X})^{-1}$ , is another option where  $tr(\mathbf{X}'\mathbf{X})^{-1}$  provides the total variance of all effects in the assumed model.

Two design criteria that focus on quantifying the transition of bias will be studied in Chapter 4 of this dissertation as we consider protection against model misspecification due to *underfitting* or *overfitting* the true system response model. As an illustration, consider model (2.1) partitioned as

$$\mathbf{y} = \beta_0 \mathbf{1} + \mathbf{X}_1 \boldsymbol{\beta}_1 + \mathbf{X}_2 \boldsymbol{\beta}_2 + \boldsymbol{\epsilon} \quad (2.21)$$

where  $\mathbf{X}_1$  and  $\mathbf{X}_2$  are  $n \times p_1$  and  $n \times p_2$  matrices respectively. If we leave out  $\mathbf{X}_2 \boldsymbol{\beta}_2$  when  $\boldsymbol{\beta}_2 \neq 0$ , then we are underfitting. On the other hand, if we include  $\mathbf{X}_2 \boldsymbol{\beta}_2$  when  $\boldsymbol{\beta}_2 = 0$ , then we are overfitting. Problems and tradeoffs result from such misspecifications of the true model. Namely, underfitting provides precise but biased estimates while overfitting provides unbiased but less precise estimates. The bias due to underfitting can be shown by the following derivation. Consider the expectation for our assumed true model coefficient estimates  $\hat{\boldsymbol{\beta}}_1$  using the usual least squares equation (2.8), we then have

$$\begin{aligned} E[\hat{\boldsymbol{\beta}}_1] &= E[(\mathbf{X}_1' \mathbf{X}_1)^{-1} \mathbf{X}_1' \mathbf{y}] \\ &= (\mathbf{X}_1' \mathbf{X}_1)^{-1} \mathbf{X}_1' [\mathbf{X}_1 \boldsymbol{\beta}_1 + \mathbf{X}_2 \boldsymbol{\beta}_2] \\ &= \boldsymbol{\beta}_1 + (\mathbf{X}_1' \mathbf{X}_1)^{-1} \mathbf{X}_1' \mathbf{X}_2 \boldsymbol{\beta}_2 \\ &= \boldsymbol{\beta}_1 + \mathbf{A} \boldsymbol{\beta}_2. \end{aligned}$$

Therefore, the bias of the least squares estimate of  $\boldsymbol{\beta}_1$  would be

$$E[\hat{\boldsymbol{\beta}}_1] - \boldsymbol{\beta}_1 = \mathbf{A} \boldsymbol{\beta}_2 \quad (2.22)$$

where  $\mathbf{A} = (\mathbf{X}_1' \mathbf{X}_1)^{-1} \mathbf{X}_1' \mathbf{X}_2$  is known as the *alias* matrix, and can be viewed as responsible for transmitting bias errors to the estimate  $\hat{\boldsymbol{\beta}}_1$ . From this result, Bursztyn and Steinberg (2006) proposed the minimization of  $tr(\mathbf{A} \mathbf{A}')$  as a way of minimizing the possible impact of this bias since the values of  $\boldsymbol{\beta}_2$  are not known *a priori* and so cannot be fully assumed to follow a multivariate normal distribution. In Myers et al. (2009, pg. 505) it is recognized that if the model misspecification is substantial, then the quality of our fitted values and thus our estimate of residual mean square



$$s^2 = \frac{(\mathbf{y} - \mathbf{X}_1 \hat{\boldsymbol{\beta}}_1)' (\mathbf{y} - \mathbf{X}_1 \mathbf{b} \hat{\boldsymbol{\beta}}_1)}{n - k - 1} \quad (2.23)$$

in the analysis of variance can also be biased since

$$\begin{aligned} E[\hat{\mathbf{y}}] &= \mathbf{X}_1 E[\hat{\boldsymbol{\beta}}_1] \\ &= \mathbf{X}_1 (\boldsymbol{\beta}_1 + \mathbf{A} \boldsymbol{\beta}_2) \\ &= \mathbf{X}_1 \boldsymbol{\beta}_1 + \mathbf{X}_1 \mathbf{A} \boldsymbol{\beta}_2 \end{aligned}$$

and

$$\begin{aligned} E[\hat{\mathbf{y}}] - E[\mathbf{y}] &= [\mathbf{X}_1 \boldsymbol{\beta}_1 + \mathbf{X}_1 \mathbf{A} \boldsymbol{\beta}_2] - [\mathbf{X}_1 \boldsymbol{\beta}_1 + \mathbf{X}_2 \boldsymbol{\beta}_2] \\ &= [\mathbf{X}_1 \mathbf{A} - \mathbf{X}_2] \boldsymbol{\beta}_2. \end{aligned} \quad (2.24)$$

Thus, the expectation of our system variance  $\sigma^2$  can be stated as

$$\begin{aligned} E[s^2] &= \sigma^2 + \frac{1}{p_1} \{\boldsymbol{\beta}_2' [\mathbf{X}_1 \mathbf{A} - \mathbf{X}_2]' [\mathbf{X}_1 \mathbf{A} - \mathbf{X}_2] \boldsymbol{\beta}_2\} \\ &= \sigma^2 + \frac{1}{p_1} \{\boldsymbol{\beta}_2' \mathbf{R}' \mathbf{R} \boldsymbol{\beta}_2\}, \end{aligned} \quad (2.25)$$

and therefore the bias of the estimate for  $\sigma^2$  as

$$E[s^2] - \sigma^2 = \frac{1}{p_1} \{\boldsymbol{\beta}_2' \mathbf{R}' \mathbf{R} \boldsymbol{\beta}_2\} \quad (2.26)$$

where the matrix  $\mathbf{R} = \mathbf{X}_1 \mathbf{A} - \mathbf{X}_2$ . Because of this result, Myers et al. suggest using the minimization of  $tr(\mathbf{R}' \mathbf{R})$  as a design criterion can help reduce the possible impact of positive bias being transmitted to the estimate of  $\sigma^2$  and consequently lowering the power to detect active effects.

## 2.5 Simulation Setting

As an extension of the steps shown below, the simulation used to investigate the two-stage sequential approach is provided in Chapter 5. We now present the following steps to detail the algorithm used to simulate the different one-stage initial screening approaches investigated in this dissertation.

For each design being considered, in each of  $i$  iterations:

1. Letting  $\mathbf{X}$  be the design matrix,  $m$  columns of  $\mathbf{X}$  are randomly assigned as active main effects.
2. For  $t = 1$  to 7,  $t$  two-factor interaction columns are selected from  $\mathbf{X}$  based upon the  $m$  columns chosen from step 1 and using weak effect heredity.
3. A sample set of possible coefficient values  $\boldsymbol{\beta}_{active} = \{\beta_1, \beta_2, \dots, \beta_{m+t}\}$  is generated by sampling from an exponential distribution with  $\mu = 4.5$ , truncated with lower and upper bounds of 1 and 10 respectively.  $m$  of the  $\boldsymbol{\beta}_{active}$  values are then randomly assigned as coefficient values for the  $m$  active main effects determined by step 1. Coefficient values for the  $t$  active two-factor interactions determined in step 2 are then obtained by sampling from a subset of  $\boldsymbol{\beta}_{active}$ , where none of the values in this subset are greater than the minimum  $\beta$  assigned to any of the  $m$  main effects. A sign of  $\pm$  is then randomly assigned to each of the  $m + t$  coefficients.
4. Letting  $\mathbf{X}_a$  be the matrix consisting of the  $m$  and  $t$  columns of  $\mathbf{X}$ , a single response vector is formed as  $\mathbf{y} = \mathbf{X}_a \boldsymbol{\beta} + \boldsymbol{\epsilon}$ , where  $\boldsymbol{\epsilon} \sim N(0,1)$ .
5. Analysis is conducted using one of the following four procedures:
  - FS restricted by weak effect heredity and EER=0.5 controlled by Bonferroni adjusted  $p$ -values when using a model-dependent error
  - FS restricted by weak effect heredity and EER=0.5 controlled by Bonferroni adjusted  $p$ -values when using a model-independent error
  - the Dantzig selector with  $\gamma = 1$  and  $\delta^*$  determined by  $BIC$
  - the Dantzig selector with  $\gamma = 1$  and  $\delta_{pe} = \hat{\sigma}_{PE} \sqrt{2 \log(p)}$

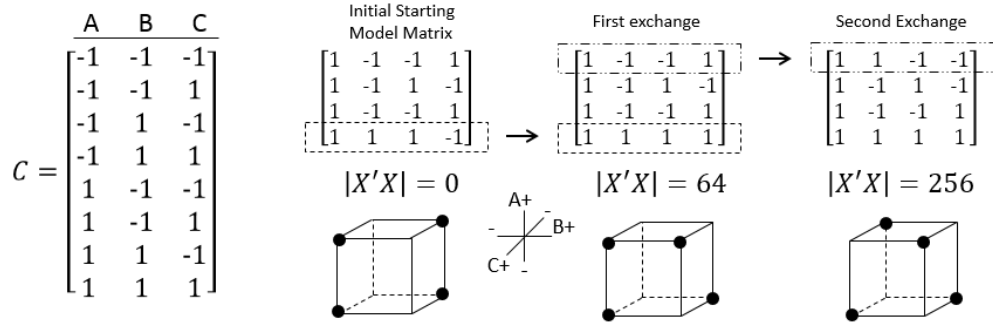
All simulation results implementing forward selection are from 2,000 iterations. Simulation results implementing the Dantzig selector when model-independent and model-dependent error estimates are used are from 1,000 and 500 iterations respectively due to the increase in computation time. Mean coverage probabilities, false discovery rates (FDR's), power, and type I error rates are calculated separately for main-effects and two-factor

interactions and since screening experiments focus on the ability to detect the most active effects (whether main-effects or interactions), we also present results for different effect sizes (i.e., dominate, moderate, and small). After analysis, estimated effect sizes are categorized in magnitude as dominant ( $|\beta| > 7$ ), moderate ( $2 < |\beta| \leq 7$ ), or small ( $1 < |\beta| \leq 2$ ). Categories were chosen such that each category and their possible combinations would allow for variability in model complexity. Inspection of the complete set of simulated responses indicates that more than 25% of the simulated truth models contained at least one dominate effect, more than 80% contained at least one small effect, and that more than 95% contained at least one moderate effect. Due to weak effect heredity, we note that more dominant effects are simulated for models when more of the main effects are simulated as active.

## **2.6 Design Algorithms**

### **2.6.1 Candidate Exchange Algorithm**

The construction of an optimum design with respect to a chosen criterion can be viewed as an optimization problem that can be approached with different methods. One such method is a candidate-exchange algorithm (Fedorov, 1972). This type of algorithm iteratively exchanges a design point (i.e., a row of a design matrix or experimental run) from some initial starting design for another design point belonging to a candidate set of possible points deemed permissible by the user. Each iteration checks for improvement in the objective function as defined by the user specified criterion. This iterative process terminates once row exchanges no longer improve the objective function.



*Figure 2.3 Candidate-exchange algorithm example.*

To illustrate this procedure, a simple example is provided in Figure 2.3. Suppose we are trying to fit a main-effects only model in three factors {A, B, and C} using  $n = 4$  experimental runs. Also suppose, we decide that the candidate set, C, of possible choices is a full factorial in these three factors. In this example, the initial starting design has an objective function of  $|X'X| = 0$  since columns for factors B and C are completely correlated with each other. When this occurs we say that the design is singular and the inverse of  $X'X$  does not exist. However, we know that a positive  $D$ -optimal criterion value can be found as long as  $n > p$ . By exchanging the last row, the algorithm increases the objective function to 64. A further exchange of the first row produces an additional improvement from 64 to 256 where the procedure terminates since no other row exchange will result in improvement. In order to help ensure a global rather than local optimum design is constructed, the process is usually repeated with a new random starting design or random start of the initial rows to be exchanged if a fixed set of design points are desired in the final design. In this example, we know we have an optimal result since  $X'X$  for the final resulting design is a diagonal matrix. However, this procedure can result in multiple optimum designs for the same criterion. Later in Chapter 4, the use of multiple criteria will be shown to help guide similar construction

algorithms in order to further improve the combined characteristics of an experimental design.

## 2.6.2 Coordinate-Exchange Algorithm

The *DP*-optimal designs created in Gilmour and Trinca (2012) all use a candidate exchange algorithm; however, other types of exchange algorithms that aim to optimize an objective function also exist such as the coordinate-exchange algorithm (Meyer and Nachtsheim, 1995), so named because as it proceeds element by element through the rows of the design matrix it is essentially proceeding element by element through the design space. For continuous factors, the algorithm begins by generating random values on the interval  $[-1, +1]$  for every element of the design matrix. Then for two-level designs, element by element, the algorithm compares objective function values when changing the element to -1 or +1 and finalizes the choice based upon the largest improvement in the objective function.

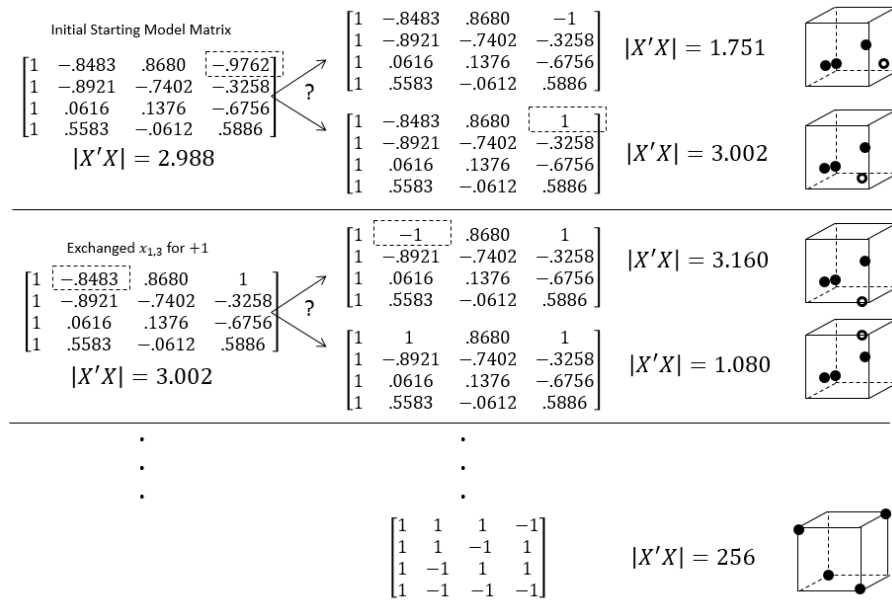


Figure 2.4 Coordinate-exchange algorithm example.

Figure 2.4 is provided to illustrate the first couple of iterations of such an approach. As in the previous example, suppose we are interested in fitting a main-effects only model for three factors using four runs. Implementing the  $D$ -optimality criterion, the randomly created initial design has an objective function of  $|X'X| = 2.9875$ . Because we are searching a design space with fixed boundaries  $[-1, +1]$  for each element, we no longer require a candidate set of design points from which to choose. Instead, we can directly begin comparing values of the objective function when changing the first randomly chosen element of the first row of the initial starting design to either -1 or +1. Changing -0.9762 to -1 results in  $|X'X| = 1.7510$  while changing to +1 results in  $|X'X| = 3.0024$ . Therefore, the entry is changed to +1 and the algorithm continues to compare changes to the remaining elements in the row and then proceeds to exchange elements in a similar fashion for the remaining rows of the design. The procedure is then repeated over all rows before converging upon the final solution design, where  $|X'X| = 256$ . We know this result is  $D$ -optimal since  $X'X$  is a diagonal matrix and thus contains globally maximal information, but this solution is not unique.

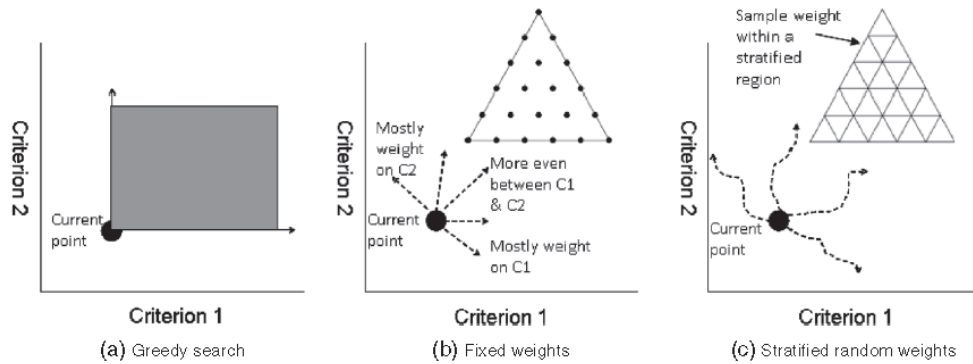
One major advantage of the coordinate-exchange algorithm is that it converges in polynomial time since it does not require a candidate set of possible design points from which to choose. Though for three factors at two levels the candidate set as a full factorial only amounts to  $2^3 = 8$  possible rows, twenty-five factors results in  $2^{25} > 33$  million rows, which can form a bottle-neck in a computer simulation routine if determinates or other computationally intensive calculations are required in the optimization objective function.

### 2.6.3 The Pareto Aggregating Point Exchange Method

The application of simultaneously optimizing multiple criteria has been a continuing development in design of experiments. In search for a similar end result as DuMouchel & Jones (1994), Allen et al. (2003) looked to simultaneously minimize parameter estimate variance and bias by developing a criterion based upon mean square error. In another approach, Jones and Nachtsheim (2011) used conditional optimization in order to optimize one criterion while attempting to preserve a threshold on another. The flexibility in optimizing one criterion seems to be greatly reduced when trying to incorporate multiple criteria into the design creation process. Attempting to regain this flexibility, Lu et al. (2011) introduced the Pareto front approach to design of experiments by developing the Pareto Aggregating Point Exchange (PAPE) method. The algorithm proceeds as follows. After randomly selecting  $n$  design points with replacement from a candidate set of design points, the method uses a candidate-exchange procedure to simultaneously optimize multiple criteria of interest. The search is free to move in the design space but in a non-detrimental fashion toward a 'utopian' design (i.e., rows are exchanged into the design only if an improvement in one or more criteria do not harm other criteria). If no such improvement is possible, the resulting design is stored and the process repeats with another randomly chosen design of run size  $n$ . Multiple random starts of this process result in a number of designs, and a front is populated by retaining only the strictly non-dominated designs.

The work of Lu et al. (2011) was later improved since the approach only made use of a greedy updating mechanism. That is, permanent row exchanges were permitted without any criterion preference. Lu and Anderson-Cook (2012) provide two alternate updating mechanisms to the greedy approach. Using a desirability function, these two enhancements

are intended to help direct the search and allow the algorithm to be more efficient. Figure 2.5 illustrates the three variations of these updating mechanisms for the two-criterion case. The greedy approach (a) is free to search the design space only in a non-detrimental direction therefore severely limiting control of the search. In the fixed weights approach (b), the search is allowed to move in directions determined by a set of weight combinations identified in a simplex that gives good coverage of the entire weighting space. In the stratified random weights approach (c), random generated weights from each strata are used with equal partitions of the weighting space to determine the direction of the search at each updating step. In Chapter 4 of this dissertation we will make use of the fixed weight approach depicted in Figure 2.5 (b) since it allows for control of the search without imposing too much limitation on the number of resulting design choices.



*Figure 2.5 Three PAPE search methods (Lu & Anderson-Cook, 2012).*



## **Chapter 3**

# **Optimum Screening Designs with and without Partial Replication**

### **3.1 Introduction**

There has been a good deal of study regarding the partial replication of screening designs (Dasgupta et al. (2010), Gilmour and Trinca (2012), Liao and Chai (2009), Lupinacci and Pigeon (2008), Mee et al. (2015), Ou et al. (2013), Tsai and Liao (2014), etc.). The reasoning for doing so stems from a desire for a reliable estimate for experimental error, which would allow the implementation of more simple analysis procedures to identify the truly active effects. Research into partial replication of designs for screening purposes is particularly important since screening methods put an emphasis on limiting the number of experimental runs to typically be less than the total number of effects to be estimated. That is, when the run size is limited to be less than the number of effects being investigated, replicating even a few of the experimental runs to produce a model-independent error estimate can reduce the ability of a screening design to perform as intended. Further, though the literature provides theoretical justification for partial replication of screening designs, the actual significance of pairing such designs with supportive analysis procedures is lacking (Mee et al. (2015), Liao and Chai (2009)). To contribute this area of research we present results from a simulation study where both forward selection and the Dantzig selector are implemented, using both a model-dependent and model-independent error estimate. 7-factor screening designs using

$n = \{20, 32\}$  runs and 10-factor screening designs using  $n = 20$  runs are used. Screening methodologies are carried out using the different combinations of designs, model selection procedures, and error-dependence strategies.

The outline of the chapter is as follows. We will first introduce some background information concerning particular screening designs which have been proposed in the literature as providing competitive screening performance. Next, we present the formulation of the Bayesian *DP*-criterion as part of a screening methodology. Simulation study results are then provided to illustrate the performance of BDP designs when compared to alternatives found in the literature. We end the chapter with some final comments and discussion.

### 3.2 *DP*-Optimality

A flexible strategy to partially replicate screening designs is to use exchange-algorithms that are driven by the optimization of a specified criterion. Gilmour and Trinca (2012) use such an approach based upon Draper and Smith (1998, pg. 144)'s formulation for the volume of a confidence interval region for  $\hat{\beta}$ ,

$$R = 2^p t^p s^p (V_{11} V_{22} \dots V_{pp})^{1/2},$$

where  $p$  is the number of parameters,  $s^2$  is an estimate of  $\sigma^2$ ,  $t$  is the  $(1 - \alpha/2)$ -quantial of the  $t$ -distribution, and  $(V_{11} V_{22} \dots V_{pp})$  are the diagonal elements of  $(\mathbf{X}'\mathbf{X})^{-1}$  since all parameter estimates are assumed uncorrelated. Substituting an  $F$  variable for the  $t$  variable and no longer assuming the rows of  $\mathbf{X}$  are uncorrelated, Gilmour and Trinca (2012) introduce the *DP*-optimality criterion, which minimizes

$$(F_{p, d:1-\alpha})^p / |\mathbf{X}'\mathbf{X}|, \tag{3.1}$$

where  $p$  is the number of parameters in the assumed model,  $d$  is the pure error degrees of freedom, and  $F_{p,d;1-\alpha}$  is the  $(1 - \alpha)$ -quantile of the  $F$ -distribution. They use a candidate-exchange algorithm and start with a set of  $n$  random design points rather than augment a fixed set of design points with replicates, therefore allowing a design of any run size to be formed as long as the assumed model is estimable. However, designs based upon (3.1) tend to have a high number of replicated runs, and therefore appear to be an inefficient solution when used to create screening designs when two-factor interactions are also considered. To investigate this assumption, we use (3.1) to create two 7-factor  $DP$ -optimal designs with  $n = \{20, 32\}$  and a 10-factor,  $n = 20$   $DP$ -optimal design and include them in the simulation study.

## 3.3 A Bayesian Framework

### 3.3.1 Bayesian $DP$ -Optimality

A drawback to the  $D$ -optimality approach is the required assumption that the assumed model is known. Model-robust approaches have been developed to allow designs to be created that yield reasonable results for the true model even if it has not been correctly postulated. To help  $D$ -optimal designs overcome dependency on an assumed model, DuMouchel and Jones (1994) introduced a simple Bayesian modification which involves the defining of *primary* and *potential* terms of interest. In this way,  $\mathbf{X}$  can be chosen such that the primary terms can be precisely estimated while still providing some estimability of the potential terms. Further, the Bayesian  $D$ -optimal approach circumvents the singularity problem of  $|\mathbf{X}'\mathbf{X}|$  in situations where  $n < p$  by adding prior information.

By definition, the primary terms are assumed active; however, the directions of these effects are unknown. Therefore, the coefficients of primary terms are given an arbitrary prior mean and a prior variance tending toward infinity. Potential terms are assumed to be inactive, and therefore given a prior mean of zero and finite variance  $\tau^2\sigma^2$ , where a choice of  $\tau = 1$  corresponds to the further assumption that the effect of any potential term is not expected to be much greater than the residual standard error. Let the prior distribution of the unknown model parameters be  $\boldsymbol{\beta}|\sigma^2 \sim N(\boldsymbol{\beta}_0, \sigma^2 \mathbf{R}^{-1})$ , where  $\mathbf{R} = \mathbf{K}/\tau^2$  is a prior covariance matrix and

$$\mathbf{K} = \begin{pmatrix} \mathbf{0}_{p_1 \times p_1} & \mathbf{0}_{p_1 \times p_2} \\ \mathbf{0}_{p_2 \times p_1} & \mathbf{I}_{p_2 \times p_2} \end{pmatrix} \quad (3.2)$$

is a matrix indicating the  $p_1$  primary terms and  $p_2$  potential terms, and assume the conditional distribution of  $\mathbf{y}$  given  $\boldsymbol{\beta}$  and  $\sigma^2$  as  $\mathbf{y} | (\boldsymbol{\beta}, \sigma^2) \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I})$ . The addition of this prior information changes the variance of estimates to  $Var(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X} + \mathbf{R})^{-1}$ , hence the Bayesian  $D$ -optimality criterion can be stated as maximizing  $|\mathbf{X}'\mathbf{X} + \mathbf{R}|$  and yields the posterior distribution for  $\boldsymbol{\beta}$  given  $\mathbf{y}$  as

$$\boldsymbol{\beta} | \mathbf{y} \sim N[(\mathbf{X}'\mathbf{X} + \mathbf{R})^{-1}(\mathbf{X}'\mathbf{y} + \mathbf{R}\boldsymbol{\beta}_0), \sigma^2(\mathbf{X}'\mathbf{X} + \mathbf{R})^{-1}]. \quad (3.3)$$

Note that  $p_1 + p_2 = p$  and that  $p_1 \leq n$  is required in order for the model to still be estimable; however, the  $p_2$  potential terms can now help influence the optimality criterion to drive down correlations (aliasing) between the primary and potential terms. That is, the Bayesian  $D$ -optimality criterion extends the consideration of driving down correlations between primary and potential effects rather than simply the correlations between the primary effects. A scaling value of  $\tau = 1$  is used for designs in this chapter as well in Chapter 4 since,

as suggested in Ruggoo and Vandebroek (2004), one-stage experiments are not shown to benefit from larger values.

Greyscale correlation maps illustrating the different aliasing structures of the non-replicated  $D$ -optimal design and a non-replicated Bayesian  $D$ -optimal design for seven factors and  $n = 20$  are shown in Figures 3.1 (a) and (b) respectively. The  $D$ -optimal design has been created for a main-effects only model while the Bayesian  $D$ -optimal design has been constructed for a main-effect with two-factor interactions model, where main-effects have been designated as primary terms and interactions as potential terms. The two black lines in each plot are intended to separate the correlation structure into three distinct sections. The upper left square section depicts correlations among main-effects, the larger square section in the lower right depicts correlations among two-factor interactions, and each of the two rectangular sections depicts correlations between each main-effect and two-factor interaction.

From these plots we can see how the incorporation of the Bayesian framework permits an exchange-algorithm to more evenly disperse the correlations among certain effects. This is not always a desired result since the partial aliasing can mask the dominant effects from detection by an analysis procedure. However, properly implemented, the Bayesian framework permits designs to be created that concentrate the forming of stronger correlations (darker cells) between known unimportant effects (potential terms) while distributing the smaller correlations (lighter cells) among known important effects (primary terms) as well as distributing smaller correlations between the known important and unimportant effects. Inspecting Figure 3.1 (a), we can see the  $D$ -optimal design is orthogonal in the main-effects. That is, it minimizes all correlations among the seven factors. Comparing

this structure to that shown in Figure 3.1 (b), we can see how the Bayesian  $D$ -optimal design sacrifices some of the orthogonality of the main-effects in order to reduce correlations between the main-effects and the two-factor interactions. Though non-orthogonality of main-effects is not a desirable design characteristic, reducing correlations between main-effects and two-factor interactions can be advantageous in cases where effect sizes of some two-factor interactions are quite large with respect to moderately sized or smaller active main-effects.

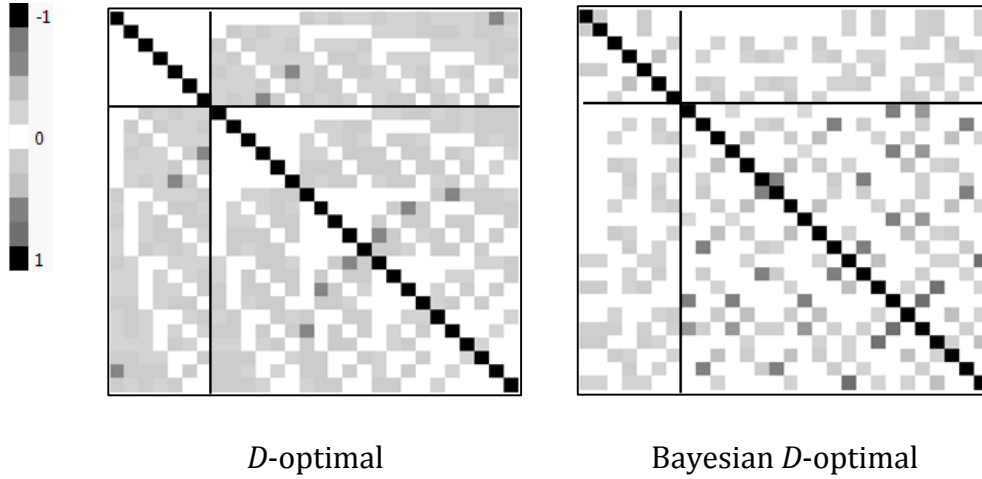


Figure 3.1 Greyscale correlation maps for 7-factor, 20-run non-replicated designs

The theoretical justification of  $D$ - and Bayesian  $D$ -optimality thus far shown assumes  $\sigma^2$  to be known; however, in practice, an estimate of  $\sigma^2$  is needed to conduct inference. When intending to use pure error for inference, Gilmour and Trinca (2012) state the importance of using design criteria, such as the  $DP$ -optimality criterion (3.1), that support the provision for a sufficient number of replicates or degrees of freedom pure error,  $df_{pe}$ , when estimating  $\sigma^2$ . Extending the formulation of the  $DP$ -optimality criterion to incorporate the Bayesian framework, we propose the Bayesian  $DP$ -optimality criterion which maximizes

$$\left| \mathbf{X}'\mathbf{X} + \frac{\mathbf{K}}{\tau^2} \right| / (F_{p,d:1-\alpha})^p. \quad (3.4)$$

where  $p = n - 1$ ,  $d$  is pure error degrees of freedom, and  $F_{p,d:1-\alpha}$  is the  $(1 - \alpha)$ -quantile of the  $F$ -distribution. We chose  $p = n - 1$  since this is the maximum number of model parameters estimable by the design. Ideally, we would use  $p = n - n_u$ , where  $n_u$  is the number of replicated design points; however, because  $n_u$  is unknown *a priori*, a value for cannot be pre-specified. Consideration could be given to  $n_u \approx 0.20p$  if effect sparsity is assumed and the number of factors being investigated results in a reasonable number of unique runs. For example, while 20 factors would result in a fairly reasonable  $n_u \approx 42$ , a consideration for 70 factors and thus  $n_u \approx 483$  is probably not.

### 3.3.2 Constructing Designs with a Fixed Number of Replicate Points

A two-stage design construction approach was used in order to allow the flexibility of forming a base design of unique points as well as the flexibility in augmenting a base design with any pre-specified fixed number of replicate points. The 2-step process first uses a coordinate-exchange algorithm to create designs based upon the Bayesian  $D$ -optimality criterion where main-effects are listed as primary terms and two-factor interactions as potential terms. Then, each designs is augmented with replicate runs using a candidate-exchange algorithm and Bayesian  $D$ -optimality to form the complete 20-run screening designs. The resulting designs, we term BDBD designs.

### 3.3.3 Comparing Properties of Designs with Varying $df_{pe}$

To compare the different designs resulting from an incorporation of each criterion (i.e., Bayesian  $DP$ -,  $DP$ -, Bayesian  $D$ -, and  $D$ -optimality) separate sets of 200 designs were created for  $k = 7$  factors,  $n = \{20, 32\}$  runs and for  $k = 10$ ,  $n = 20$  runs using each criterion. Separate sets of 200 BDBD designs of corresponding run size but each with a pre-specified fixed number  $df_{pe}$  were also created. For the 7-factor, 20-run BDBD designs, fixed  $df_{pe} = \{2, 4, 6\}$  were used; for the 7-factor, 32-run BDBD designs, fixed  $df_{pe} = \{2, 4, 6, 10, 12\}$  were used; and for the 10-factor, 20-run BDBD designs, fixed  $df_{pe} = \{2, 6, 8\}$  were used.

Using the formulations provided in Table 3.1, smoothed curves of criterion efficiencies were calculated for all created designs and are depicted in Figures 3.2, 3.3 and 3.4 for the two groups of 7-factor designs using 20 or 32 runs and the group of 10-factor designs using 20 runs.  $DP$ -efficiencies for the designs are shown in the upper left plots in each of the three figures. Because  $DP$ -optimality is an extension of  $D$ -optimality, attention is given solely to the assumed model, thereby driving the exchange-algorithm to create designs that simply fulfill the ability to estimate a main-effects only model. The result is the creation of designs which contain up to as many as  $n - k + 1$  replicate runs. However, in scenarios where a main-effect with two-factor interactions model is of interest, this results in an inefficient use of resources since it is highly unlikely in practice that none of the two-factor interactions are active.



Table 3.1 Formulations for criterion efficiencies

$DP_{eff} = \left( \frac{\{ \mathbf{X}(\xi)' \mathbf{X}(\xi) ^{\frac{1}{k+1}}\} / F_{k+1, df_{pe}}}{\{ \mathbf{X}(\xi^*)' \mathbf{X}(\xi^*) ^{\frac{1}{k+1}}\} / F_{k+1, df_{pe}^*}} \right)$	$BDP_{eff} = \left( \frac{\{ \mathbf{X}(\xi)' \mathbf{X}(\xi) + W ^{\frac{1}{n-1}}\} / F_{n-1, df_{pe}}}{\{ \mathbf{X}(\xi^*)' \mathbf{X}(\xi^*) + W ^{\frac{1}{n-1}}\} / F_{n-1, df_{pe}^*}} \right)$
$BD_{eff} = \left( \frac{ \mathbf{X}(\xi)' \mathbf{X}(\xi) + W }{ \mathbf{X}(\xi^*)' \mathbf{X}(\xi^*) + W } \right)^{\frac{1}{n-1}}$	$D_{eff} = \left( \frac{ \mathbf{X}(\xi)' \mathbf{X}(\xi) }{ \mathbf{X}(\xi^*)' \mathbf{X}(\xi^*) } \right)^{\frac{1}{k+1}}$

In the lower left plots of the three figures a somewhat opposite effect can be seen with the Bayesian  $D$ -criterion. This criterion aims to combine runs which will permit estimation of the main-effects model in addition to as many potential two-factor interaction effects allowable by the given run size. However, absent the  $F$ -statistic in its formulation, the Bayesian  $D$ -criterion gives no consideration to an estimation of pure error and so never replicates design points for any of designs being considered. In the upper right plots a compromising effect of combining the  $DP$ - and Bayesian  $D$ -criteria can be seen. Implementing the Bayesian  $DP$ -criterion allows for the consideration of estimating a model-independent error in addition to the ability to estimate additional two-factor interactions; however, in the upper right plot of Figure 3.4, a number of BDBD designs with  $df_{pe} = 6$  are shown to actually be more BDP-efficient. This may be due to the coordinate-exchange procedure used to construct the BDBD designs being more efficient in constructing optimal designs and that the candidate-exchange algorithm used in combination with the Bayesian  $DP$ -criterion simply needed more iterations in order to converge to solution design containing fewer replicate runs. Further research is planned to investigate this result.

The lower right plots indicate that overall  $D$ -efficiencies are not severely compromised when implementing the Bayesian  $DP$ -criterion. That is, a balance appears to be made of the competing requirements to estimate pure error and some potential two-factor interactions without heavily compromising the requirement to properly estimate a main-effects model. Though not depicted in the figures, BDBD designs were created with the same number of  $df_{pe}$  as the BDP designs and resulted in designs having similar  $D$ -efficiencies as the BDP designs. In particular, the slight dip in  $D_{eff}$  for the 7-factor BDP designs shown in the lower right plots of Figures 3.2 and 3.3 were also apparent when constructing BDBD with the same number of  $df_{pe}$ .

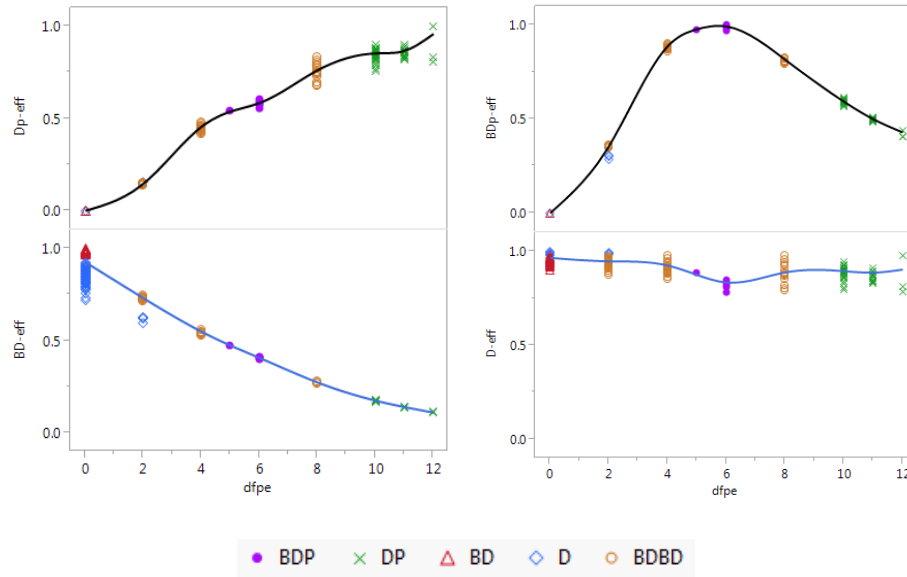


Figure 3.2 Efficiencies for 7-factor, 20-run designs

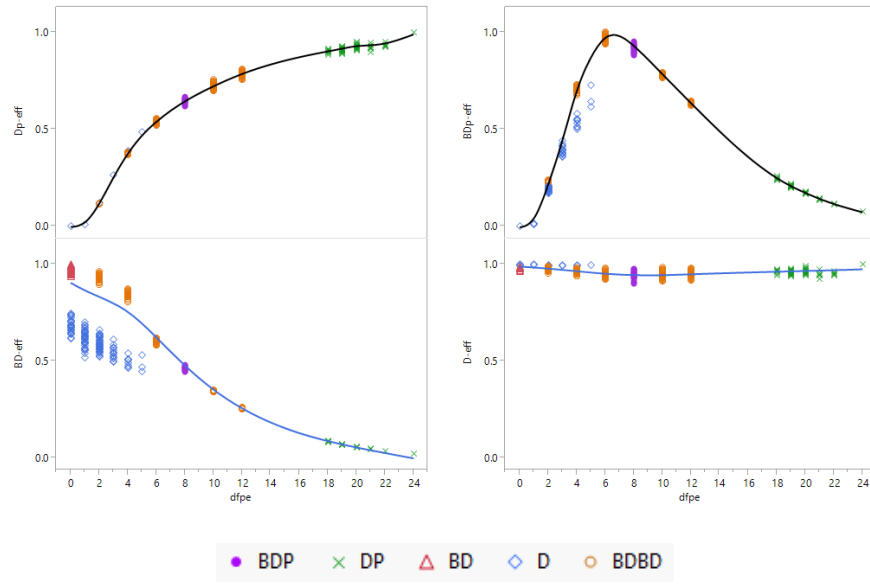


Figure 3.3 Efficiencies for 7-factor, 32-run designs

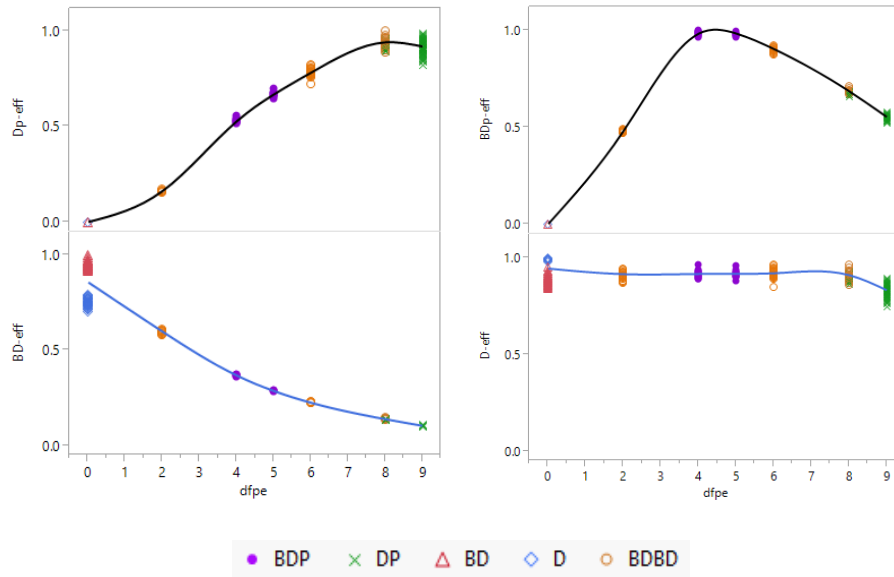
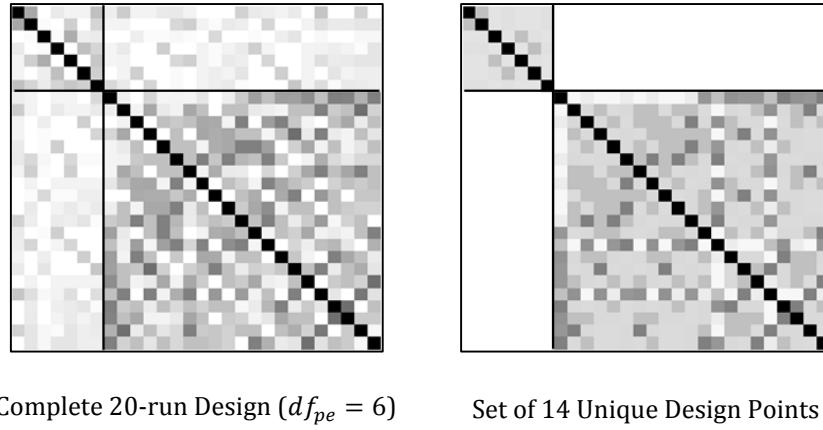


Figure 3.4 Efficiencies for 10-factor, 20-run designs

As well as reducing the ability to further decrease correlations among effects by limiting the number of unique runs, adding replicate runs to a design can potentially hinder analysis performance by increasing the transfer of bias errors to the model estimates. To

illustrate this potential impact, in Figure 3.5 we provide greyscale correlation maps comparing the aliasing structures of the complete 20-run Bayesian  $DP$ -optimal 7-factor design used in the simulation study to the set of 14 unique runs from the same design.



*Figure 3.5 Correlation maps for BDP-optimal Design and its unique runs*

We can see that the inclusion of the 6 replicate runs in the design directly compromises the orthogonality between the 7 main-effects and 28 two-factor interactions. Further, because we wish to estimate the main-effects as best as possible, our interest also lies in quantifying the amount of bias transferable to their estimators  $\hat{\beta}_1$  from the few likely active two-factor interactions absent in the assumed model. Recall (2.22) from Chapter 2

$$E[\hat{\beta}_1] - \beta_1 = A\beta_2$$

where  $A = (X_1'X_1)^{-1}X_1'X_2$  and  $tr(AA')$  proposed by Bursztyn and Steinberg (2006) as a way of quantifying the bias transferring to  $\hat{\beta}_1$  if the true model has been underfit due to  $\beta_2$  (i.e., some of the two-factor interactions which are active) not being included in the assumed truth model.

Figure 3.6 displays scatterplots of  $tr(\mathbf{AA}')$  values versus the number of replicate runs for 7-factor designs using 20 or 32 runs. The steep increase in  $tr(\mathbf{AA}')$  is clearly visible for designs with 20 runs when more than six design points are replicated. This shows that the Bayesian  $DP$ -criterion provides a means to create partially replicated designs that limit the transmission of bias to the estimates of main-effects models if active two-factor interaction terms have been unintentionally left out of the assumed model. Also shown in Figure 3.6, the same increase is noticeable for the 7-factor 32-run designs once more than eight design points are replicated, suggesting that the Bayesian  $DP$ -criterion may provide an informal upper bound on the number of replicate runs used in screening designs; however, as Figure 3.7 illustrates, the same pattern is not found when comparing the 10-factor designs with 20 runs. Instead, the increase in  $tr(\mathbf{AA}')$  is more linear for designs containing between two to eight replicate points, with a sharp increase not taking place until nine of the 20 runs are replicated. This may suggest that  $n = 20$  is too small of a run size for 10-factor screening designs when consideration is given to all two-factor interactions being potentially active.

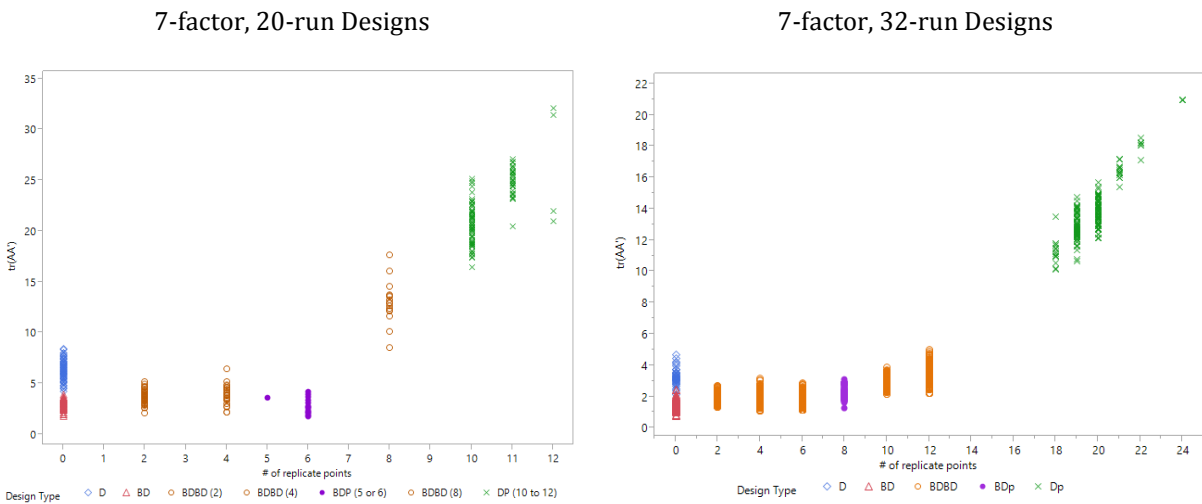


Figure 3.6 Scatterplots of  $tr(\mathbf{AA}')$  versus number of replicates for 7-factor designs

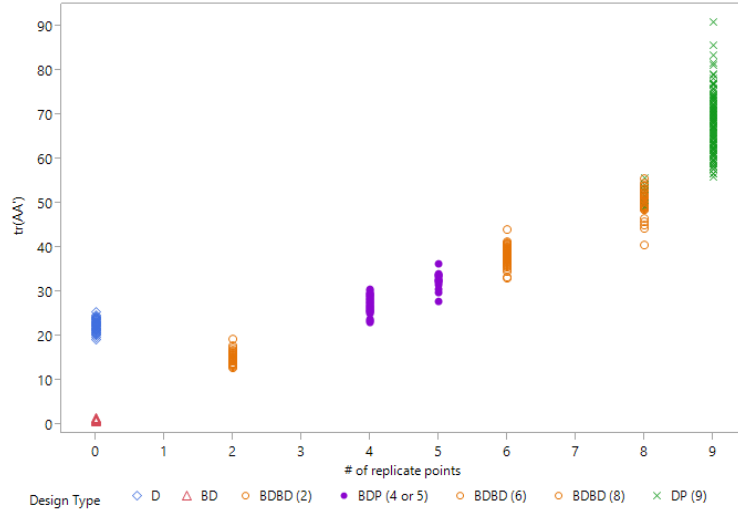


Figure 3.7 Scatterplot of  $tr(\mathbf{AA}')$  versus number of replicates for 10-factor designs

### 3.4 Simulation Results

For the three simulation studies presented in this chapter we compare equally sized designs taken as subsets from the collections of 7-factor designs with  $n = \{20, 32\}$  experimental runs and 10-factor designs with  $n = 20$  experimental runs. Designs based upon the  $D$ -, Bayesian  $D$ -, Bayesian  $DP$ -, or  $DP$ -criterion are optimal for the stated criterion unless otherwise noted (the  $k=7, n=32$  BDBD design with  $df_{pe} = 6$  is actually BDP-optimal). In cases where multiple designs were optimal, ties were broken by choosing the design with the lowest  $tr(\mathbf{AA}')$  value. That is, the optimal design considered to have the lowest impact on biasing the estimation of main-effects is chosen. The BDBD designs are chosen based upon a combination of having a high  $D_{eff}$  and low  $tr(\mathbf{AA}')$  value. Details concerning the subsets of designs are provided in each subsection before commenting on the simulation results. None of the  $D$ -optimal nor Bayesian  $D$ -optimal designs contain replicate points. Therefore, when pairing either of these designs with the Dantzig selector or forward selection only a model-

dependent error estimate will be considered. For the partially replicated designs, analyses are carried out implementing both model selection procedures using a model-independent error estimate.

### 3.4.1 Results for 7-factor designs with $n = 20$ runs

For the first simulation study presented in this chapter we compare six designs taken as a subset from the suite of 7-factor designs with  $n = 20$  experimental runs. Details concerning these six designs are provided in Table 3.2. The *DP*-optimal and Bayesian *DP*-optimal designs contain 12 and 6 replicate runs, respectively. The two BDBD designs used in this study contain 2 or 4 replicate runs in order to investigate the performance of designs containing fewer replicate points than the BDP design, which contains 6 replicates. The *DP*-optimal design contains 12 replicates and therefore leaves no *df* for estimating any of the 21 possible two-factor interactions when a main-effects only model is considered, while the unreplicated *D*-optimal and Bayesian *D*-optimal designs allow estimation of models containing up to 12 of the 21 two-factor interactions in addition to all seven main effects. The *D*-efficiencies are relatively high except for the BDP design, which is around 10% less efficient in precisely estimating a main-effects only model than the other design being considered. The Bayesian *D*-optimal design has the lowest  $tr(\mathbf{AA}')$  value, indicating that this design has the lowest potential for transmitting bias from potentially active two-factor interactions to the estimation of main effects, and the *DP*-optimal design with  $tr(\mathbf{AA}') = 21$  shows signs of possibly contributing the most bias.

Table 3.2 Characteristics of 7-factor 20-run designs

Design Type	$df_{pe}$	$D_{eff}$	$tr(\mathbf{AA}')$	# $df$ free for 2FI's
D	0	optimal	5.16	12
BD	0	0.96	1.98	12
BDBD	2	0.98	2.59	10
BDBD	4	0.98	3.88	8
BDP	6	0.87	2.69	6
DP	12	0.98	21.00	0

Coverage probabilities and false discovery rates (FDR's) for dominant, moderate, and small effect sizes are displayed in Figures 3.8 and 3.9 when 2/7 and 4/7 main effects are simulated as active, respectively. Figure 3.10 provides the mean number of main effects and two-factor interactions included in the final model. Power, type I error rates, and FDR's for main effects and two-factor interactions are illustrated in Figures 3.11 and 3.12.

- Comparing overall coverages for when two or four of the seven main effects are active shows that pairing the Dantzig selector with either of the un-replicated designs performs similarly to using forward selection when using a BDBD design with 2 or 4 replicate runs (Figures 3.8 and 3.9, row 1). Coverages for dominant effects appear highest when implementing the former approach, and FDR's are also minimized (Figures 3.8 and 3.9, row 2). However, it should be noted that fewer than half of the simulated models contain at least one dominant sized effect. Therefore, overall coverages will be less representative of an ability to detect dominant effects.
- Further comparison of using the Dantzig selector with an un-replicated design to using forward selection with a partially replicated design shows similar power to detect active main effects (Figure 3.11, row 1); however, the Dantzig selector performs slightly better with respect to detecting two-factor interactions when



paired with the un-replicated  $D$ -optimal design (Figure 3.11, row 3), where overfitting is better controlled (Figure 3.10, row 2) as well as FDR's kept lower for two-factor interactions (Figures 3.12, row 2).

- Comparing performances of the two un-replicated designs, we see that the  $D$ -optimal design rather than the Bayesian  $D$ -optimal design provides higher coverage for moderate effect sizes, especially when there are a large number of active two-factor interactions (Figures 3.8 and 3.9, row 3); however, coverages for moderate effects are similar when four of the seven effects are active. Referring back to the correlation maps in Figure 3.1, we can see that there are a few high correlations present between certain two-factor interactions in the BDP design, which may be contributing to this result.
- In a general comparison of the partially replicated designs, we can see a noticeable difference in how the two model selection procedures perform with respect to the incorporation of a model-independent error estimate. While the Dantzig selector generally performs the same, forward selection is clearly shown to benefit from the use of pure error, where coverages for all effect sizes are increased (Figures 3.8 and 3.9, rows 2, 3 and 4) while FDR's are relatively maintained (rows 6, 7, and 8).
- Comparing the two BDBD designs and the BDP design we see that coverages for dominant and moderate effects when using forward selection appear to increase as the number of replicate designs points also increase and more focus is on the assumed main-effects only model (Figures 3.8 and 3.9, rows 2 and 3); however, this is at the expense of higher FDR's (rows 6 and 7), which can be more easily seen by comparing FDR's for two-factor interactions (Figure 3.12, row 2).

- The BDBD designs appear to slightly overfit the number of main-effects when paired with forward selection and only a couple of the main effects are active, while underfitting of main-effects occurs when using the Dantzig selector (Figure 3.8, row 1). This appears to worsen for the Dantzig selector as more main effects become active; however, forward selection tends to fit the correct number of main effects as the number of active main effects increases.
- Generally, FDR's are low for dominant and moderate effect sizes and closer inspection of type I errors for main effects (Figure 3.9, row 2) show that the BDP design permits forward selection a slightly better control over incorrectly identifying main effects than when using either of the BDBD designs. This can be partially attributed to the focus of the criterion to properly estimate the main-effects model, while simultaneously protecting against model-misspecification by keeping  $tr(\mathbf{A}\mathbf{A}')$  to a minimum. Though the type I errors appear to be high, for the worst case result, this only translates into 1 out the 3 main effects being incorrectly identified as active when only 2 main effects are actually active. Further, from inspection of the FDR's for the effect sizes in Figure 3.6, these errors appear to be attributed to main effects with estimated coefficients that are small in magnitude, where  $\hat{\beta} \leq 2$ . Therefore, the result of the BDBD designs overfitting more than the BDP design is not a major concern in this case.
- The *DP*-optimal design is clearly the poorest performing. Similar to the *D*-optimality criterion, the *DP*-criterion is based upon the main-effects only model as the assumed true model; however, the *DP*-optimality criterion is also driven by a criterion which continues to reduce as additional replicate points are added to the design. With a

focus on a main-effects only model and having only enough  $df$  to estimate a model with eight terms, the  $DP$ -optimal design doesn't permit analysis to consider models with interaction terms. This is clearly visible in Figure 3.8, where the mean number of main-effects and two-interaction terms contained in the final model are illustrated. The results indicate that the  $DP$ -optimal design paired with forward selection results in an overfitting of main effects (row 1), especially when only two of the seven main effects are active. However, when paired with the Dantzig selector, the  $DP$ -optimal design consistently shows no terms being included in the model even though some terms have been simulated to be almost ten times the size of  $\sigma$ .

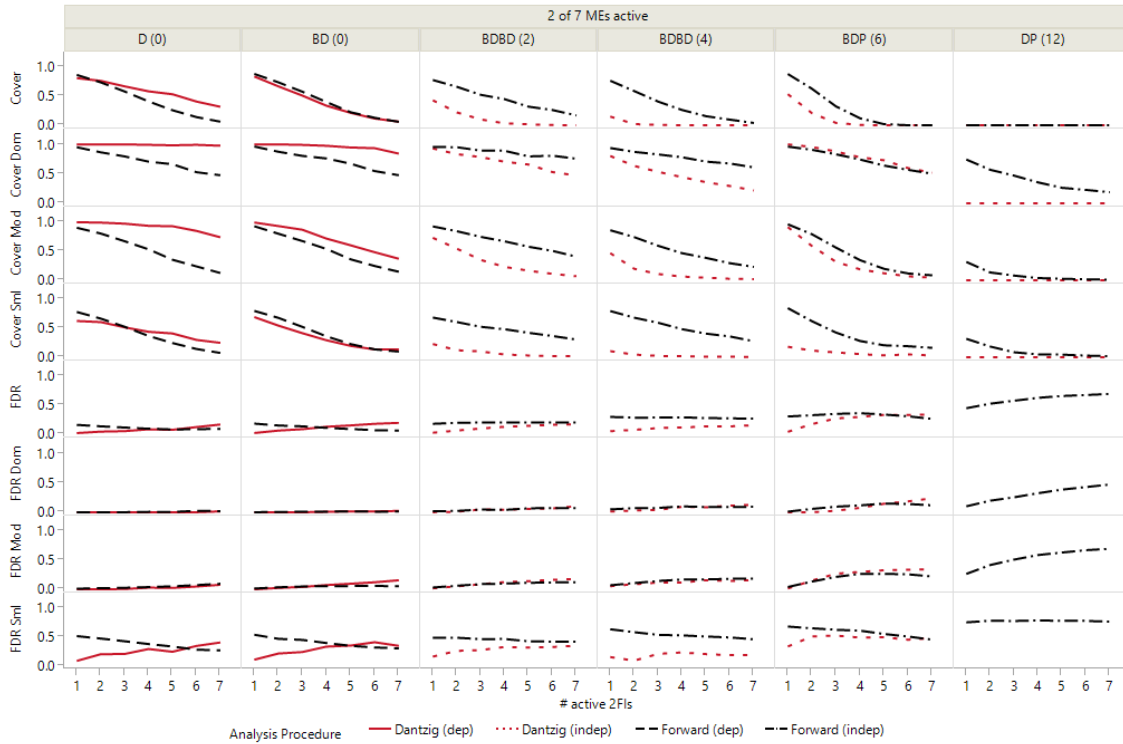


Figure 3.8 Cover and FDR's when 2/7 main effects are active ( $n=20$ )

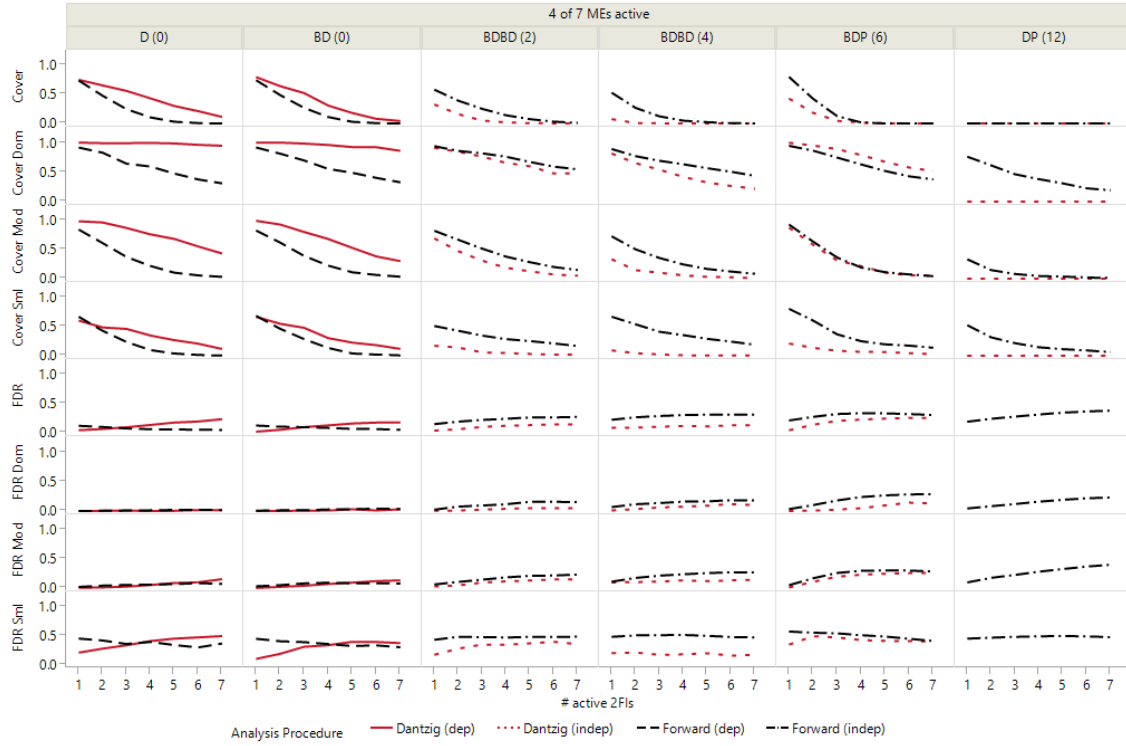


Figure 3.9 Cover and FDR's when 4/7 main effects are active ( $n=20$ )

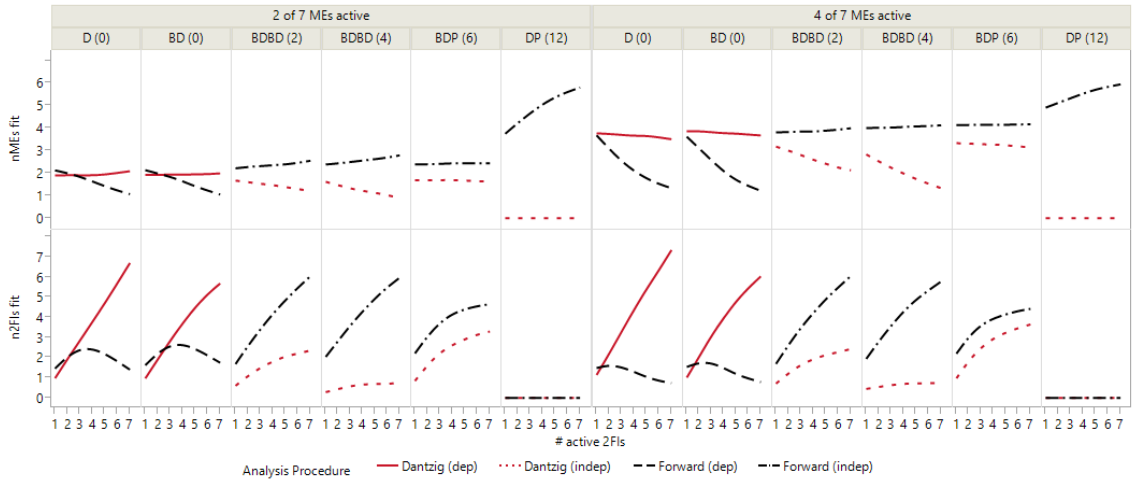


Figure 3.10 Number of terms in final model ( $k=7$ ,  $n=20$ )

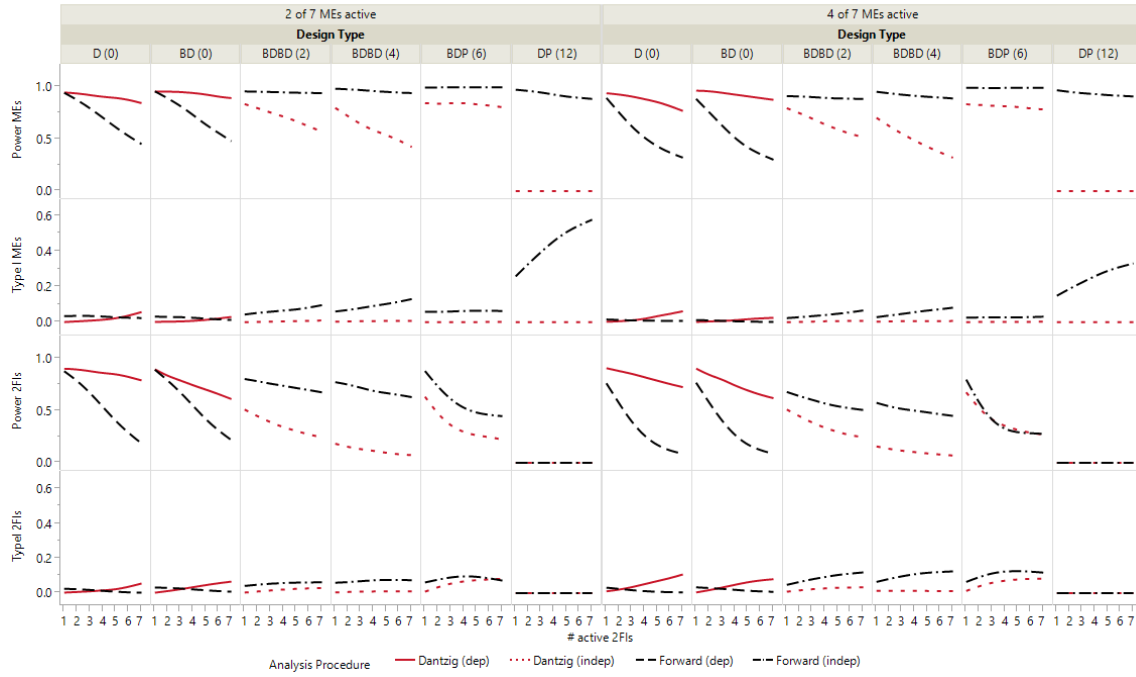


Figure 3.11 Power and type I error rates for effects ( $k=7, n=20$ )

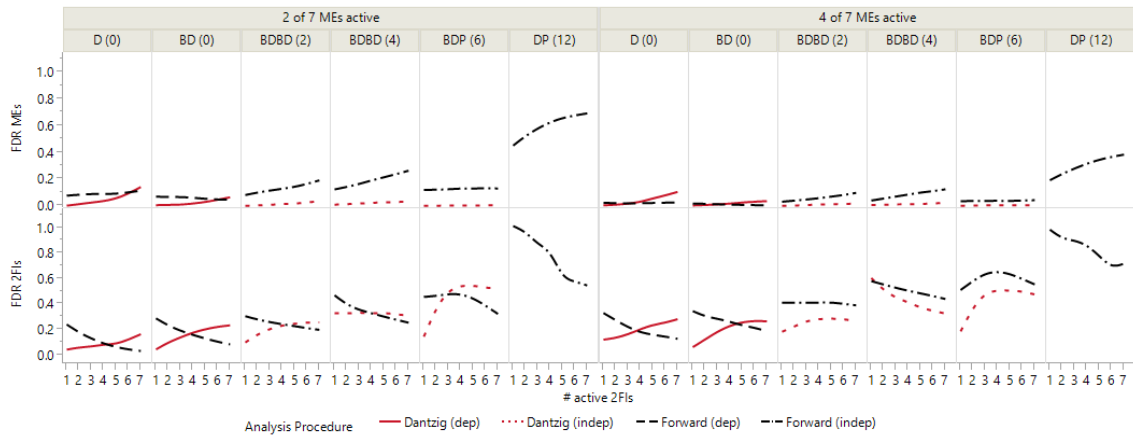


Figure 3.12 FDR's for main effects and two-factor interactions ( $k=7, n=20$ )

### 3.4.2 Results for 7-factor designs with $n = 32$ runs

For the second simulation study presented in this chapter seven designs are taken as a subset from the suite of 7-factor designs with  $n = 32$  experimental runs. Details concerning these seven designs are provided in Table 3.3. The  $DP$ -optimal and Bayesian  $DP$ -optimal designs contain 24 and 8 replicate runs, respectively. Three BDBD designs are used in this study. Two of the three BDBD designs contain two or four replicate runs in order to investigate the performance of designs containing fewer replicate points than the BDP design. A third BDBD design with 12 replicate runs is also used to identify performance characteristics when more replicates are used than contained in the BDP design. The  $DP$ -optimal design contains 24 replicates and therefore leaves no  $df$  for estimating two-factor interactions when a main-effects only model is considered, while the un-replicated  $D$ -optimal and Bayesian  $D$ -optimal designs allow estimation of models containing all 21 two-factor interactions in addition to all seven main effects. The  $D$ -efficiencies are relatively high for all designs and the  $DP$ -optimal design is also  $D$ -optimal. The BDBD design containing four replicate points has the lowest  $tr(\mathbf{AA}')$  value of 1.08, the  $D$ -optimal design has a slightly higher value of 4.18, and the  $DP$ -optimal design with  $tr(\mathbf{AA}') = 21$  indicates having the highest potential to transmit bias to the estimation of main effects when active two-factor interactions are not included in the model.

Table 3.3 Characteristics of 7-factor 32-run designs

Design Type	$df_{pe}$	$D_{eff}$	$tr(\mathbf{AA}')$	# $df$ free for 2FI's
D	0	optimal	4.18	24
BD	0	0.98	1.15	24
BDBD	2	0.97	1.34	22
BDBD	4	0.96	1.08	20
BDP	8	0.96	1.94	16
BDBD	12	0.96	2.43	12
DP	24	optimal	21.00	0

Figures of results are provided in the same format as the first study with coverage probabilities and false discovery rates (FDR's) for dominant, moderate, and small effect sizes displayed in Figures 3.13 and 3.14 when 2/7 and 4/7 main effects are simulated as active, respectively. Figure 3.15 provides the mean number of main effects and two-factor interactions included in the final model. Power, type I error rates, and FDR's for main effects and two-factor interactions are illustrated in Figures 3.16 and 3.17.

- Understandably, by increasing the run size from 20 to 32, overall performance is greatly improved compared to the first simulation. Coverage is maximized and FDR's minimized for dominant and moderate sized effects and when using either of the unreplicated designs with the Dantzig selector or when using either of the BDBD designs containing two or four replicates with forward selection.
- Forward selection struggles to detect all the active effects when paired with an unreplicated design (Figures 3.13 and 3.14), due to the lack of a model-independent error estimate. This combination appears to provide the ability to detect the correct number of main effects when 2/7 are active (Figure 3.15, row 1); however, performance deteriorates as the number of active effects increases. This can be seen

in Figure 3.16, where power to detect active effects decreases considerable with respect to other combinations of designs and selection procedures being investigated (rows 1 and 2).

- As the amount of replication increases (moving left to right in Figures 3.13 and 3.14) performance begins to noticeably deteriorate once eight design points are replicated, especially when considering the detection of moderately sized effects (row 3).
- Overall coverage is maximized using the BDBD design with two replicates paired with forward selection (Figures 3.13 and 3.14, row 1). FDR's for dominant and moderate sized effects are also minimized (rows 6 and 7). This is due, in part, to the model-independent error estimate aiding in analysis. The result also stems from the design having 30 unique design points, which reduce the partial aliasing of effects and also provide enough  $df$  to estimate all possible main-effect and two-factor interaction models. Similar performance with forward selection can also be seen when using the BDBD design with four replicate points. FDR's for small effect sizes are positive for both designs (row 8) and help explain the type I error rates for main effects and two-factor interactions (Figure 3.16) when using forward selection.
- Pairing either of the BDBD designs containing two or four replicates with the Dantzig selector also provides maximum coverage for dominant and moderate effect sizes (Figures 3.13 and 3.14, rows 2 and 3). The overall reduction in coverage (row 1) can be attributed to the selection procedure's inability to detect small effects (row 4); however, type I error rates are minimized for both main effects and two-factor interactions (Figure 3.16) and the number of terms included in the model appear to be more on target than when using forward selection (Figure 3.15). Closer inspection



of Figure 3.15 also indicates that the Dantzig selector slightly underfits the number of two-factor interactions, while forward selection overfits both the number of main effects and two-factor interactions. However, as already explained, these additional terms falsely detected by forward selection are estimated as being small effects (i.e.,  $\hat{\beta} < 2$ ).

- Similarly sized models appear to be formed by either selection procedure when paired with the BDP design (Figure 3.15); however, coverage for moderate effects is no longer maximized and FDR's are slightly positive (Figures 3.13 and 3.14, rows 3 and 7). Inspecting FDR's of main effects and two-factor interactions in Figure 3.17 indicates that the Dantzig selector has a better ability to minimize the false detection of main effects when paired with the BDP design and any of the BDBD designs, including the BDBD design with 12 replicates.
- The *DP*-optimal design performs the poorest when paired with either model selection procedure. With forward selection, the mean number of main effects included in the model is overfit (Figure 3.15, row 1) and no two-factor interactions are included (row 2). Similar to the results from the first study presented in this chapter, using the Dantzig selector in combination with the *DP*-optimal design results in a null model likely to be chosen.

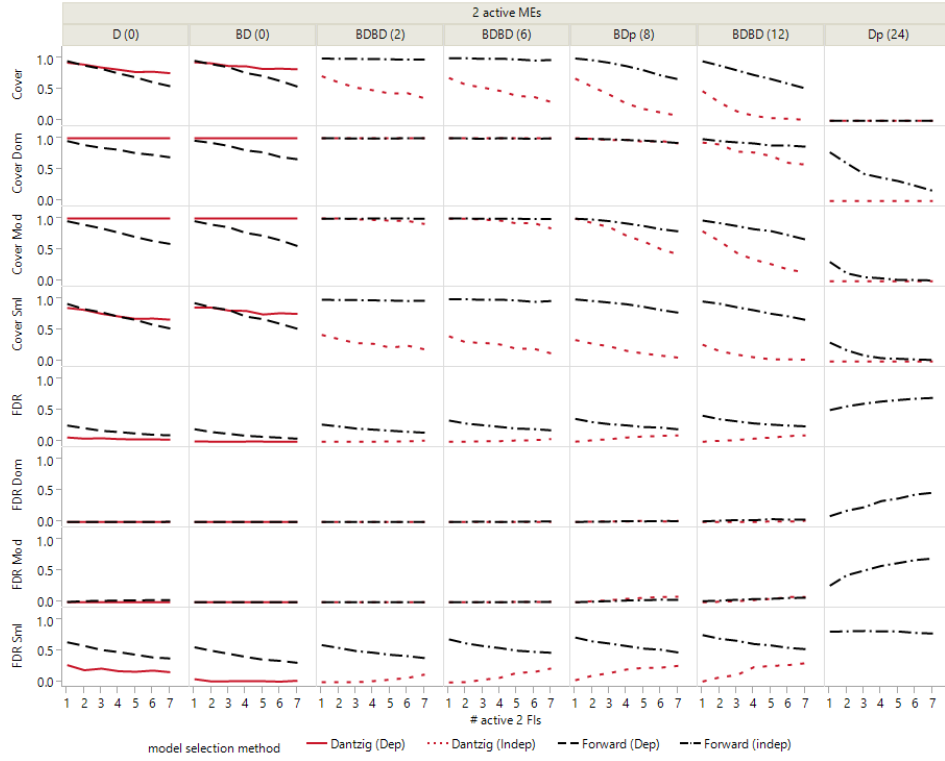


Figure 3.13 Cover and FDR's when 2/7 main effects are active ( $n=32$ )

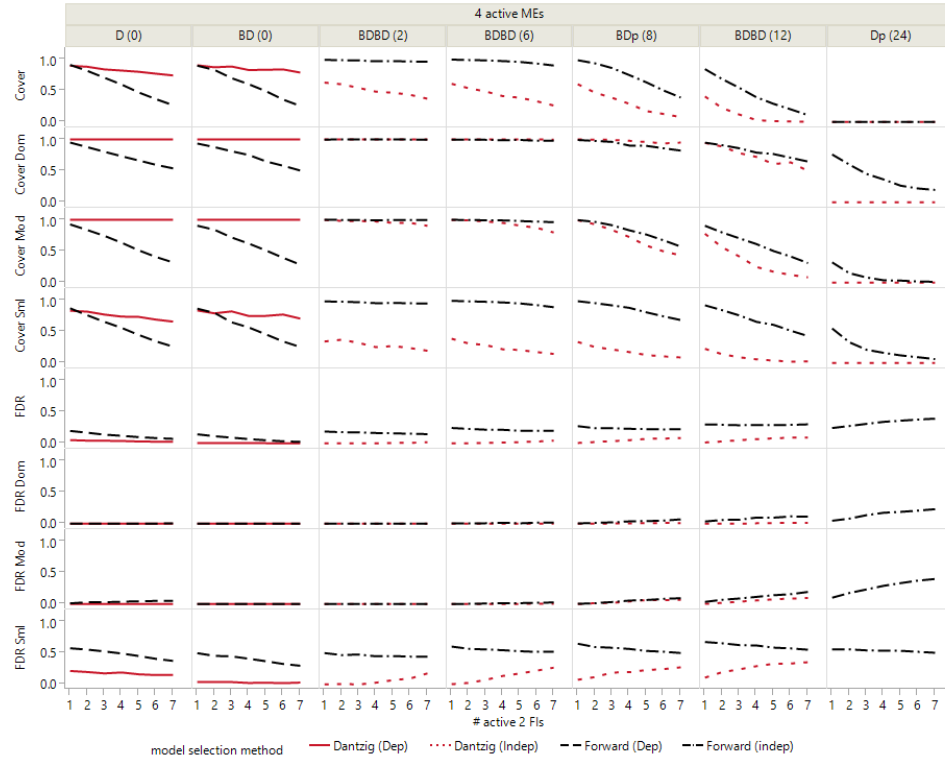


Figure 3.14 Cover and FDR's when 4/7 main effects are active ( $n=32$ )

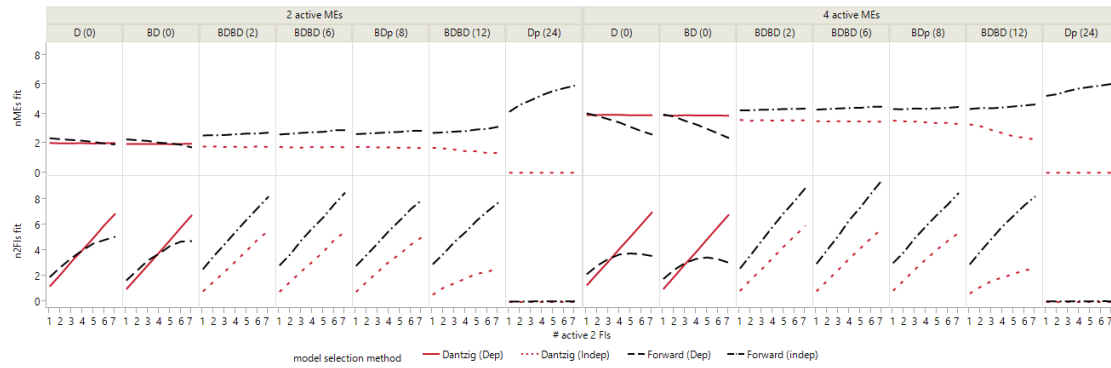


Figure 3.15 Number of terms in final model ( $k=7, n=32$ )

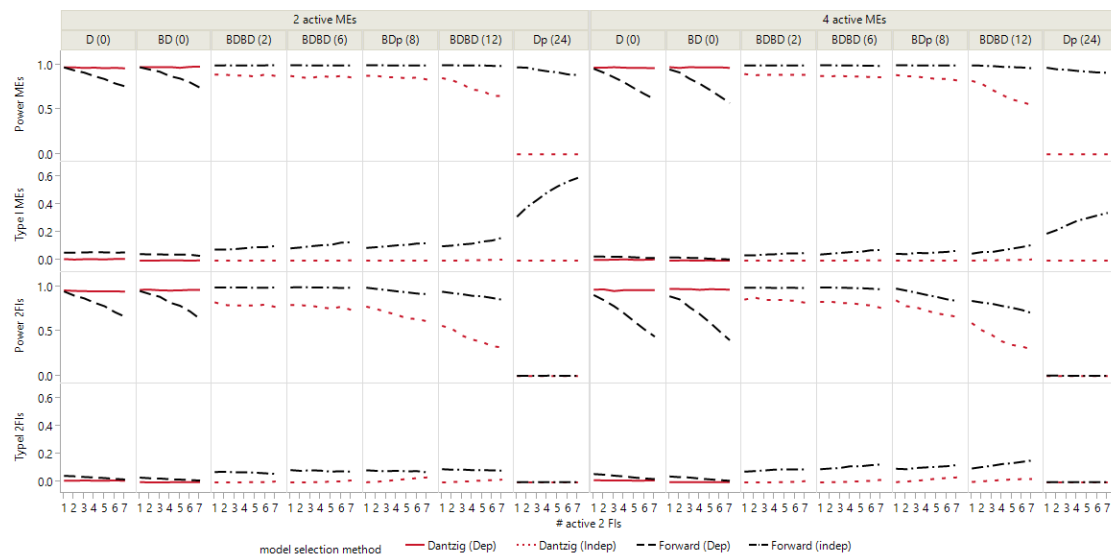
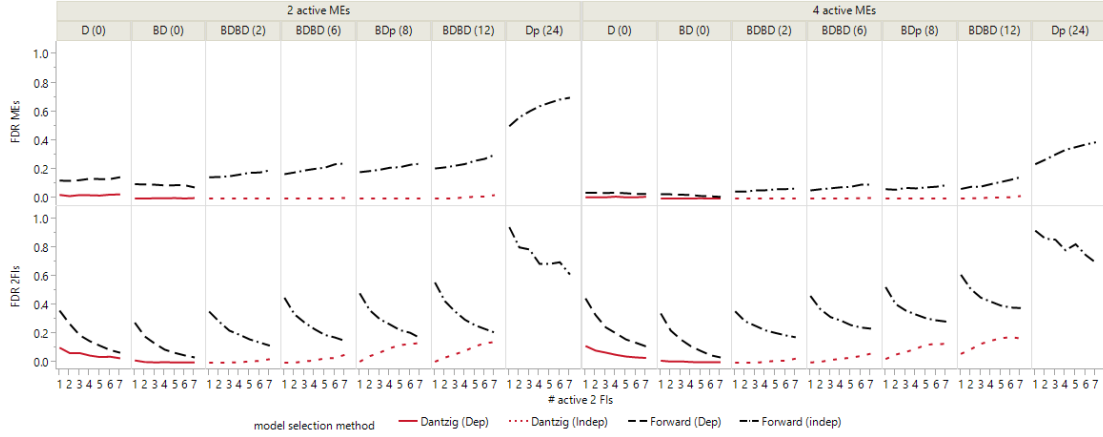


Figure 3.16 Power and type I error rates for effects ( $k=7, n=32$ )



3.17 FDR's for main effects and two-factor interactions ( $k=7, n=32$ )

### 3.4.3 Results for 10-factor designs with $n = 20$ runs

For the third simulation study presented in this chapter six designs are taken as a subset from the suite of 10-factor designs with  $n = 20$  experimental runs. Details concerning these six designs are provided in Table 3.4. The  $DP$ -optimal and Bayesian  $DP$ -optimal designs contain nine and four replicate runs, respectively. Two BD&D designs are used in this study and contain two or six replicate runs in order to investigate the performance of designs containing more or fewer replicate points than the BDP design. Having nine replicate points, the  $DP$ -optimal design leaves no  $df$  free for estimating two-factor interactions when a main-effects only model is considered, while the un-replicated most  $D$ -efficient and Bayesian  $D$ -optimal designs allow estimation of models containing nine of the possible 55 two-factor interactions in addition to all ten main effects. The Bayesian  $D$ -optimal design and BD&D design with two replicates have the lowest  $D$ -efficiencies of 0.91 and 0.94, respectively. The Bayesian  $D$ -optimal design minimizes  $tr(\mathbf{A}\mathbf{A}')$ , and therefore allows each of the 10 main effects to be estimated independently from any two-factor interactions possibly included in

a model; however, there is partial aliasing among the main effects ( $\rho = 0.20$ ) and some two-factor interactions are fully confounded ( $\rho = 1$ ) as shown in the correlation map of the design in Figure 3.18. The correlation map of the most  $D$ -efficient design used in this study is also depicted in Figure 3.18. For this design, there are only two pairs of partially aliased main effects, namely  $x_2$  with  $x_4$  and  $x_9$  with  $x_{10}$  (both with  $\rho = 0.2$ ). The BDBD design with two replicates has the next lowest value of 12.84, and similar to the two previous studies in this chapter the  $DP$ -optimal design has the highest potential to transmit bias with  $tr(\mathbf{AA}') = 55.94$ .

Table 3.4 Characteristics of 10-factor 20-run designs

Design Type	$df_{pe}$	$D_{eff}$	$tr(\mathbf{AA}')$	# $df$ free for 2FI's
D	0	0.99	22.12	9
BD	0	0.91	0	9
BDBD	2	0.94	12.84	7
BDP	4	0.97	23.10	5
BDBD	6	0.97	33.13	3
DP	9	0.97	55.94	0

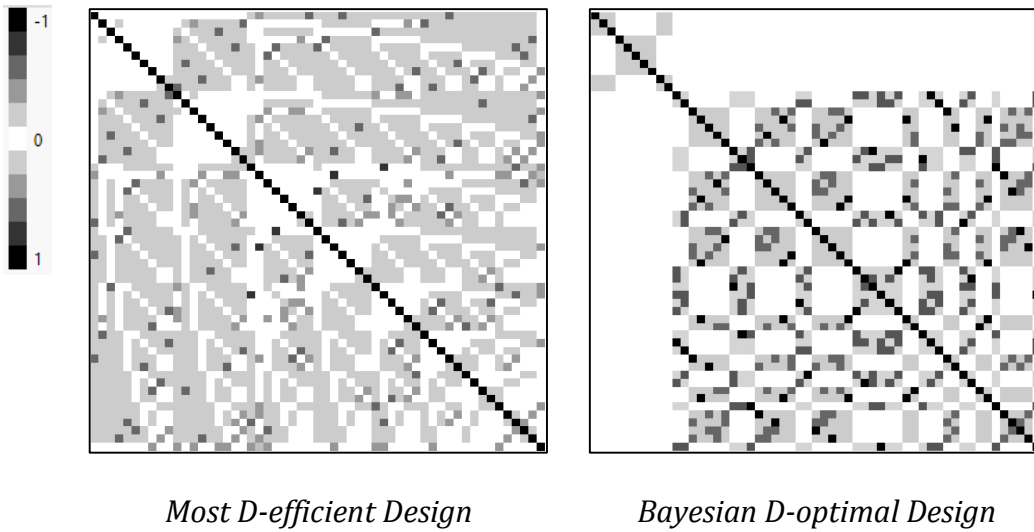


Figure 3.18 Correlation map of 10-factor Bayesian  $D$ -optimal design ( $n=20$ )

Figures of results are provided in the same format as the previous two studies with coverage probabilities and false discovery rates (FDR's) for dominant, moderate, and small effect sizes displayed in Figures 3.19 and 3.20 when 2/10 and 8/10 main effects are simulated as active, respectively. Figure 3.21 provides the mean number of main effects and two-factor interactions included in the final model. Power, type I error rates, and FDR's for main effects and two-factor interactions are illustrated in Figures 3.22 and 3.23.

- Overall performance is poor for this size of design. When 8/10 main effects are active, all effects with moderate to small effect sizes are rarely if ever detected (Figure 3.20, rows 2 and 3) and the false identification of inactive effects as dominant effects is only minimized when using forward selection with the most *D*-efficient design (row 6). Although many of the combinations result in models containing the correct number of main effects and two-factor interactions when 2/10 main effects are active, all results indicate a severe underfitting of active main effects when 8/10 main effects are active. This can be partially attributed to the distribution of active coefficients used in the simulation, and further research is needed to expand the types of responses used in simulation.
- The most *D*-efficient design paired with the Dantzig selector provides the highest coverage for dominant and moderate effects (Figures 3.19 and 3.20, rows 2 and 3); however, the combination allows for the false identification of inactive effects as dominant effects when 8/10 main effects are active (Figure 3.20, row 5) and could be associated with either main effects or two-factor interaction terms (Figure 3.23).

- With a  $D_{eff} = 0.91$  and a minimization of  $tr(\mathbf{A}\mathbf{A}')$ , the Bayesian  $D$ -optimal design shows slightly higher power to detect active main effects than the most  $D$ -efficient design when paired with either selection procedure (Figure 3.22, row 1) and with lower false detection rates (Figures 3.19 and 3.20 rows 5-8, and Figure 3.23 row 1); however, only when paired with forward selection does the design permit the detection of two-factor interactions (Figure 3.22, row 3, and Figure 3.21, row 2).
- Power to detect active main effects tends to be highest when using the BDBD design with two replicate points or the BDP design combined with forward selection; however this is at the expense of higher type I errors (Figure 3.22, rows 1 and 2). Further, these combinations result in positive FDR's for all effect sizes, indicating that inactive effects are often incorrectly estimated as being dominant effects in some cases.
- Whereas in the previous two simulations, the  $DP$ -optimal design combined with the Dantzig selector resulted in no two-factor interaction being detected and in some instances a null model, here the combination appears to form models with more two-factor interactions than main effects (Figure 3.21).

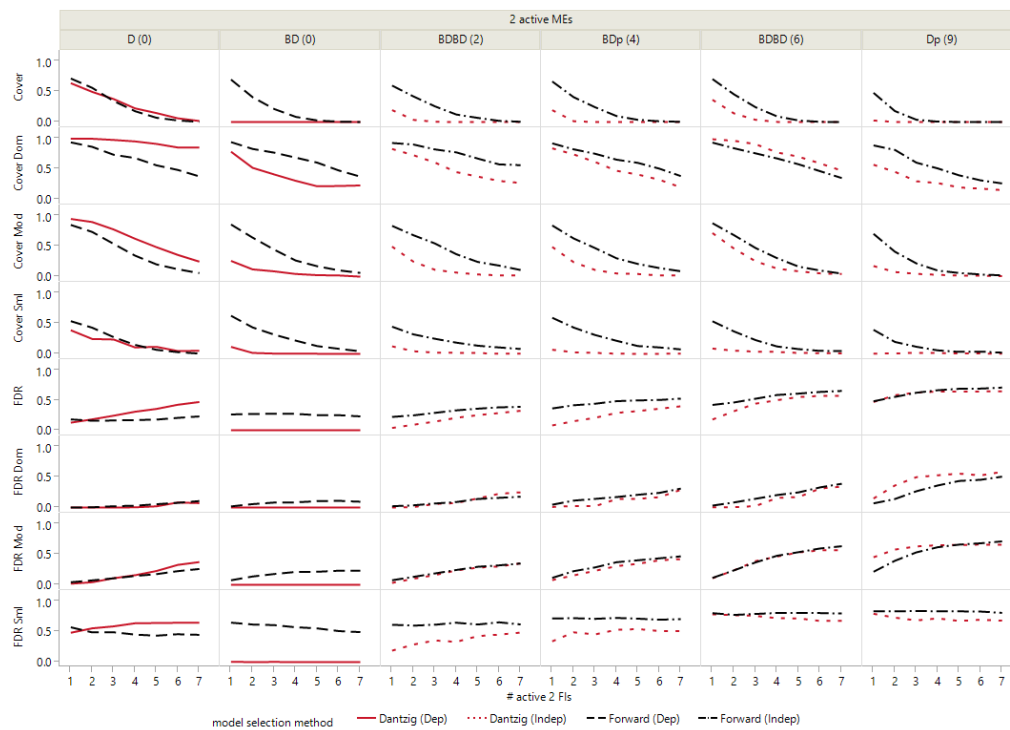


Figure 3.19 Cover and FDR's when 2/10 main effects are active ( $n=20$ )

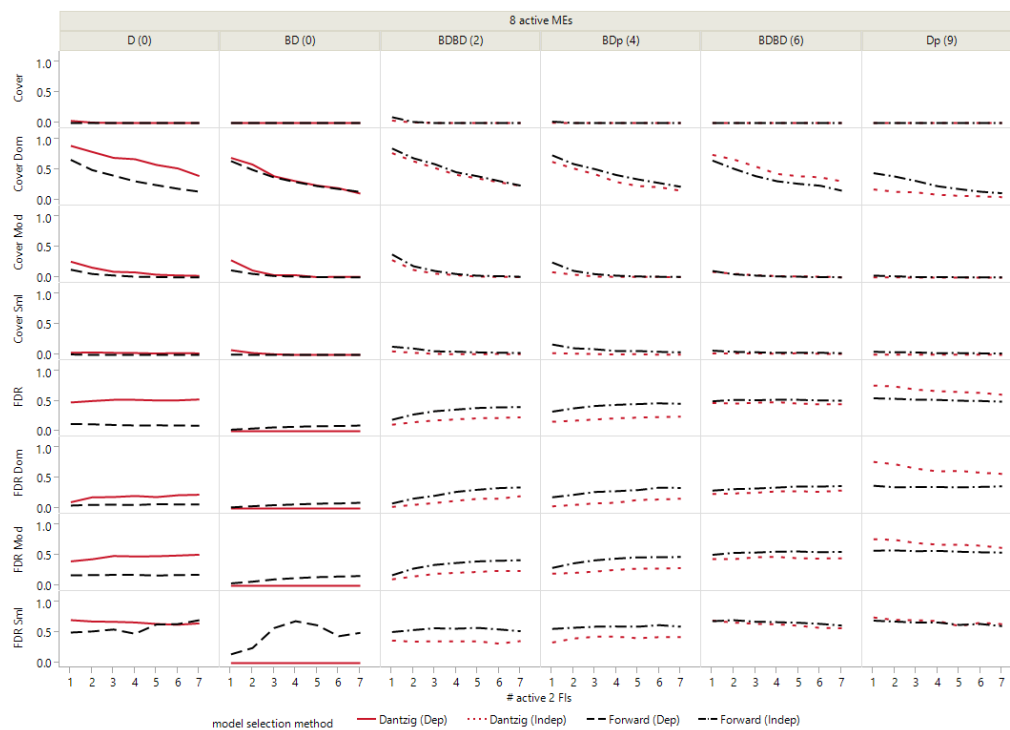


Figure 3.20 Cover and FDR's when 8/10 main effects are active ( $n=20$ )



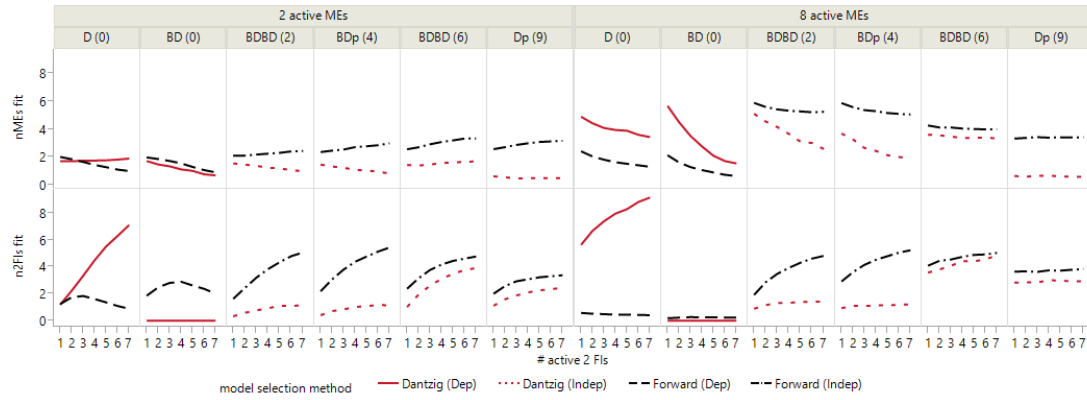


Figure 3.21 Number of terms in final model ( $k=10, n=20$ )

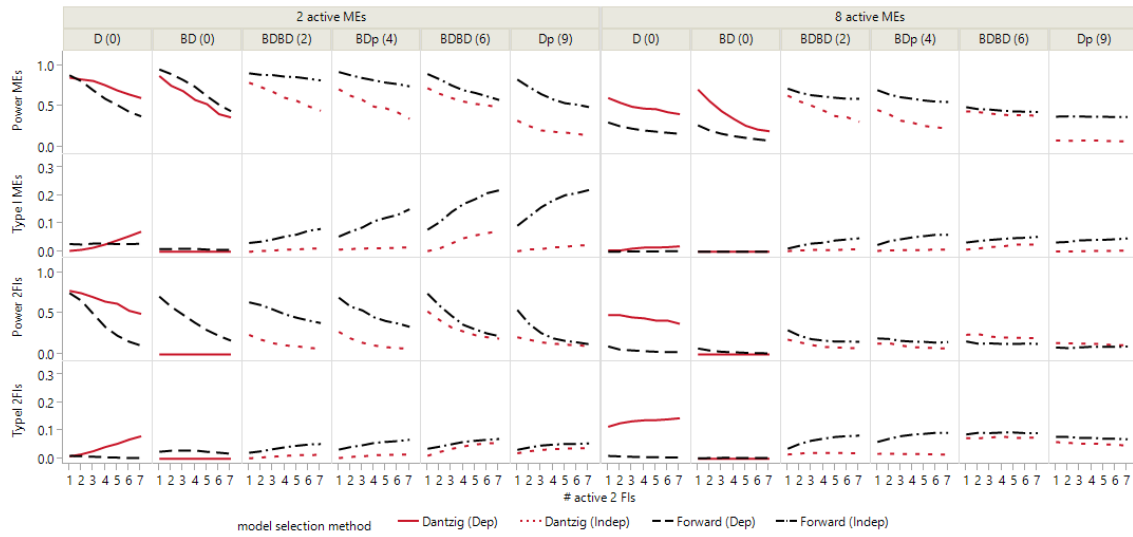
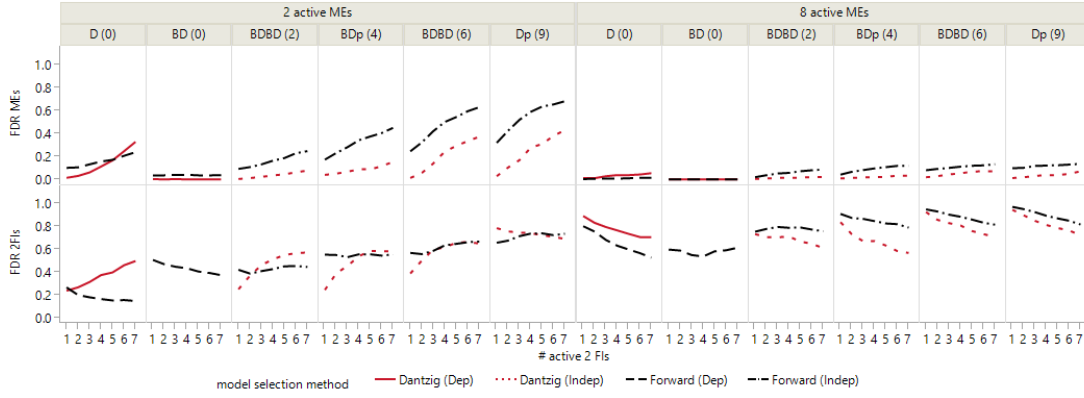


Figure 3.22 Power and type I error rates for effects ( $k=10, n=20$ )



3.23 FDR's for main effects and two-factor interactions ( $k=10, n=20$ )

### 3.5 Final Comments and Discussion

In this chapter we have proposed a Bayesian modification of the  $DP$ -criterion and illustrated some of its properties. In addition to providing flexibility in determining an optimal number of replicate designs points, the criterion has been shown to provide a possible upper bound on  $tr(\mathbf{AA}')$ , thus keeping the bias associated with non-specified model terms to a minimum. However, a number of 7-factor, 32-run designs created using a two-step construction procedure were found to be more BDP-efficient than the BDP-'optimal' design created via a candidate-exchange algorithm using the actual Bayesian  $DP$ -criterion. Further, although an abrupt change in  $tr(\mathbf{AA}')$  was apparent for the 7-factor BDP designs, investigation of 10-factor designs with  $n = 20$  runs shows the increase in  $tr(\mathbf{AA}')$  to be linear until  $df_{pe}$  equal the number of terms identified as inactive, which corresponds to the  $df_{pe}$  in the  $DP$ -optimal designs. Further, simulation results appear to show that the impact of degrading performance (i.e. lower power/coverage, higher FDR's) tends to coincide to abrupt changes in  $tr(\mathbf{AA}')$ . That is, for a given run size, there appears to be a point at which the consideration

for using unique design points to further reduce aliasing of effects may be better suited as replicates, which can be used to estimate a model-independent error for analysis purposes. However, further research is needed to determine if the Bayesian *DP*-criterion can be used for this purpose.

This chapter also involved the investigation of incorporating a model-independent error into the forward selection procedure and the Dantzig selector. Forward selection showed to benefit from the incorporation of a model-independent error estimate, with increases in power to detect active effects, an improved ability to form more properly sized models (i.e., correct number of main effects and two-factor interactions), and an increased ability to detect effects of all sizes. This was at the expense of higher FDR's and type I errors being committed, although for screening purposes may be an acceptable compromise.

Simulation results show the Dantzig selector tends to perform better when combined with un-replicated designs. In the first two simulation studies, the procedure performed similarly when combined with either the most *D*-efficient or Bayesian *D*-optimal design, where the former combination showed small improvements in coverage when two of the seven main effects were active. In the third and final study, noticeable differences emerged when pairing the procedure with both design. While both designs displayed similar ability to detect active main effects, the Bayesian *D*-optimal design failed to detect any of the active two-factor interactions. Further, combining the Bayesian *D*-optimal design with forward selection not only showed higher power to detect main effects with minimal type I error when compared to the Dantzig selector, forward selection was also shown to provide acceptable power in detecting two-factor interactions with minimal type I errors rates.

## Chapter 4

# Screening Methods & Protecting Against Model Misspecification

### 4.1 Introduction

As pointed out in Edwards and Kelly (2014), Robinson and Anderson-Cook (2011), Lu and Anderson-Cook (2012), and Lu et al. (2011), the choice of an experimental design should not be limited to a single criterion since resulting designs are not always unique and therefore further discrimination among competing designs can take place. Li and Nachtsheim (2000) provide model robust factorial designs (MRFD's) which take into account the ability to estimate model parameter estimates as well as the limitations imposed by the size of models being considered. Dasgupta et al. (2010) extended MRFD's to include optimally chosen replicated design points; however, simulations pairing either non-replicated or partially-replicated MRFD's with analysis procedures are not found in the literature. A summary of MRFD's will be provided in the first section of this chapter.

Also in this chapter, we incorporate the Bayesian  $D$ -optimality and Bayesian  $DP$ -optimality criteria into the PAPE algorithm in order to create non-replicated and partially-replicated screening designs using more than one criterion. Two additional criteria are used to help create designs which are more robust to model misspecification. Recall (2.22) from Chapter 2

$$E[\hat{\beta}_1] - \beta_1 = A\beta_2,$$

where  $\mathbf{A} = (\mathbf{X}'_1\mathbf{X}_1)^{-1}\mathbf{X}'_1\mathbf{X}_2$ , and  $tr(\mathbf{A}\mathbf{A}')$  proposed by Bursztyn and Steinberg (2006) as a way of quantifying the bias transferring to  $\hat{\boldsymbol{\beta}}_1$  if the true model has been underfit due to  $\boldsymbol{\beta}_2$  (i.e., some of the two-factor interactions which are active) not being included in the assumed truth model. Also recall from Chapter 2 that if the model misspecification is substantial, then the estimate of residual mean square (2.23) can also be biased since

$$\begin{aligned} E[\hat{\mathbf{y}}] - E[\mathbf{y}] &= [\mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_1\mathbf{A}\boldsymbol{\beta}_2] - [\mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2] \\ &= [\mathbf{X}_1\mathbf{A} - \mathbf{X}_2]\boldsymbol{\beta}_2, \end{aligned}$$

and that the bias of the estimate for the system variance  $\sigma^2$  is therefore

$$E[s^2] - \sigma^2 = \frac{1}{p_1}\{\boldsymbol{\beta}'_2\mathbf{R}'\mathbf{R}\boldsymbol{\beta}_2\}$$

where the matrix  $\mathbf{R} = \mathbf{X}_1\mathbf{A} - \mathbf{X}_2$ . Further recall, Myers et al. (2009) suggest using the minimization of  $tr(\mathbf{R}'\mathbf{R})$  as a design criterion to help reduce the possible impact of positive bias being transmitted to the estimate of  $\sigma^2$  and consequently lowering the power to detect active effects. In addition to using a multi-criterion strategy in this chapter for optimal design of experiments, the Pareto front approach to reducing the number of competing designs will be introduced with attention given to using  $df_{pe}$  as an *ad hoc* design criterion. We end the chapter by presenting simulation results comparing performances of selected designs.

## 4.2 Model-Robust Factorial Designs

Introduced by Li & Nachtsheim (2000), MRFD's have maximal estimation capacity (EC) for designs with a given number of factors and run size assuming an upper bound,  $g$ , for the number of potentially active two-factor interactions where  $g$  is less than the total number of interactions possible. Take the following example from Li & Nachtsheim (2000). Consider a

$2^{4-1}$  resolution *IV* design with defining relation  $I=1234$ . Because interest lies in estimating a first order with two-factor interactions model, in four factors there are a total of 15 possible models to consider. However, due to the complete confounding of two-factor interaction pairs  $12=34$ ,  $13=24$ , and  $14=23$ , this design can only estimate 12 of the 15 models. Therefore, the estimation capacity of this design would be  $EC=12/15=80\%$ . The authors show that if we limit our scope of potentially active two-factor interactions and focus on maximizing the information concerning only the models of this limited size (i.e., maximizing average  $D_{eff}$  of all estimable models), a design more suited to our needs could be found. Continuing the authors' example, if the number of potentially active two-factor interactions were thought to be one or two, a design of the same  $n = 8$  run size but with 100% EC for all possible  $\binom{m}{g} = \binom{4}{2} = 6$  models containing 4 main-effects and 2 two-factor interactions is found. For our simulation study, we consider the 8-factor case using the 12-run MRFD A-optimally augmented ( $4df_{pe}$ ) of Dasgupta et al. (2010) and the un-replicated 16-run MRFD of Li & Nachtsheim (2000) and compare simulated performance with designs constructed when incorporating the Bayesian *DP*-optimality criterion into the PAPE algorithm of Lu et al. (2011). Estimation capacities for the two MRFDs along with other designs being considered in this chapter are provided in section 4.4.

## 4.3 The Pareto Front Approach

### 4.3.1 The Pareto Front Approach - An Example from the Literature

Figure 4.1 illustrates the Pareto front approach as presented in Lu et al. (2011) for a general two-criterion case. The approach is meant to reduce the set of competing design choices by

eliminating from consideration the designs which are dominated in all criteria by at least one other design. Each of the five points on the Pareto front represents a strictly non-dominated design and the utopian point in the figure represents an ideal solution design with the best of both criterion values. In this example, the ideal solution design would maximize Criterion 1 and minimize Criterion 2. This ideal design is usually not attainable; however, by using the Pareto front approach, trade-offs between different candidate solution designs can be explored.

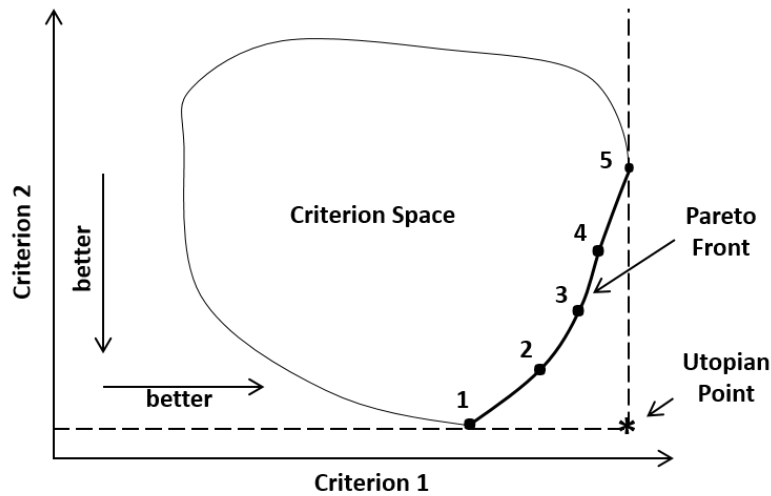


Figure 4.1 Illustration of Pareto front in two dimensions

For example, let Criterion 1 be  $D_{eff}$  and Criterion 2 be  $tr(\mathbf{AA}')$  and say the Criterion Space contains 300 possible designs. Let us assume that, as depicted in Figure 4.1, that out of the 300 possible designs only five designs reside on the Pareto front and let these designs correspond to the values given in Table 4.1. Design 5 is the most  $D$ -efficient and therefore provides the best ability to precisely estimate the assumed model; however, Design 5 also has the highest  $tr(\mathbf{AA}')$  value, indicating that the design has the most potential to transmit bias to the estimation of effects in the assumed model if some truly active effects have not

been included. This design would therefore be used in situations where there is firm belief that the assumed model is correct. On the other hand, if there is high uncertainty in the model assumption, then perhaps using Design 1, which minimizes  $tr(\mathbf{AA}')$  and thus potentially eliminates possible bias transfer, would be worth the reduction in  $D$ -efficiency. Further, if a compromise in both criterion is possible, then perhaps one of the other three designs on the Pareto front would be a better choice.

*Table 4.1  $D_{eff}$  and  $tr(\mathbf{AA}')$  values for Pareto front example*

Design	$D_{eff}$	$tr(\mathbf{AA}')$
1	0.76	0
2	0.85	6.5
3	0.90	8.2
4	0.95	12.4
5	1	19

Though the type of plot shown in Figure 4.1 aids in the general comparison of results, deciding upon a final solution could still be difficult. Lu et al. (2011) offer a solution to this problem by exploring a range of preference weightings on the combined criteria. First, each criterion is scaled to a  $[0,1]$  interval, where 0 corresponds to the worst value on the front and 1 to the best. Then, weighted scores are created using each pair of scaled criterion values corresponding to each design on the Pareto front. The design with the highest score for any given weighting scheme is recorded. The reoccurrence of a design across a wide range of different weighting schemes implies that the design is robust to uncertainty in user preference for any particular criterion. Results from performing calculations with a grid of weights increasing by one-half of a percent, results in the distribution of weights shown in Figure 4.3. Although Design 2 resides on the Pareto front, results indicate that either Design



1 or Design 3 is a comparable choice, therefore reducing the set of competing choices from five to four. Designs 1 and 5 appear to be the more robust choices, indicating that extreme preference in one criterion or the other is not necessary if deciding to use one of these two designs. Designs 3 and 4 both appear to be possible choices if uncertainty in criterion preference were high.

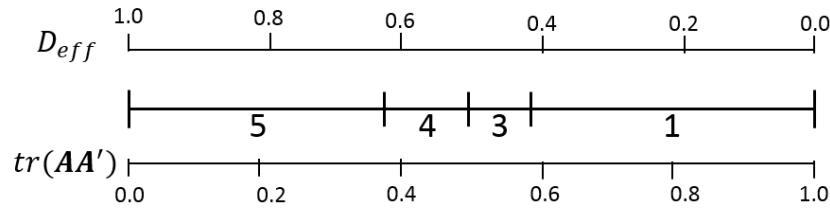


Figure 4.2 Plot indicating distribution of weightings for example designs

The PAPE algorithm can also incorporate more than two criteria. Lu et al. (2011) portray the optimization of a two-level, 14-run design to estimate a first order with two-factor interactions model using  $D_{eff}$ ,  $tr(\mathbf{AA}')$ , and  $tr(\mathbf{R}'\mathbf{R})$ . All main effects and three of the two-factor interactions were deemed important and used to calculate the three criteria values for each design. 40,000 random starts were based upon these three criteria and the pair-wise scatterplots depicting the 333 solutions (designs) are shown in Figure 4.3 (a). The black filled squares  $[D_{eff} \& tr(\mathbf{AA}')$ , diamonds  $[D_{eff} \& tr(\mathbf{R}'\mathbf{R})]$ , and triangles  $[tr(\mathbf{AA}') \& tr(\mathbf{R}'\mathbf{R})]$  denote solutions on their respective two-criterion Pareto fronts. The lighter grey points represent the remaining solutions on the three-criterion Pareto front.

Figure 4.3 (b) illustrates their approach with a mixture plot. By running normalized criteria values of all 333 designs on the three-criterion Pareto front through a fine mesh of

weights, all plausible weighting schemes can be inspected for their best result. By doing so, the authors show the 333 design choices are then reduced to a non-dominated set of 17 design choices, which makes the comparison of competing design choices more manageable. Further, the authors present the mixture plot as a way to inform the practitioner of the ranges of possible weights to consider when judging their own uncertainty about the impact of each criterion on their design choice. For example, notice in Figure 4.3 (b) design 1 appears fairly robust to most weighting, and therefore probably most attractive if the practitioner were uncertain in valuing each criterion. However, design 4 might be chosen if a balance between  $D_{eff}$  and  $tr(\mathbf{A}\mathbf{A}')$  were desired but  $tr(\mathbf{R}'\mathbf{R})$  was of little concern. That is, if precise model parameter estimates and protection against model misspecification contributing to random error variance were of importance but the potential for model misspecification contributing to systematic error bias were not much of a concern, then design 4 would be a good choice.

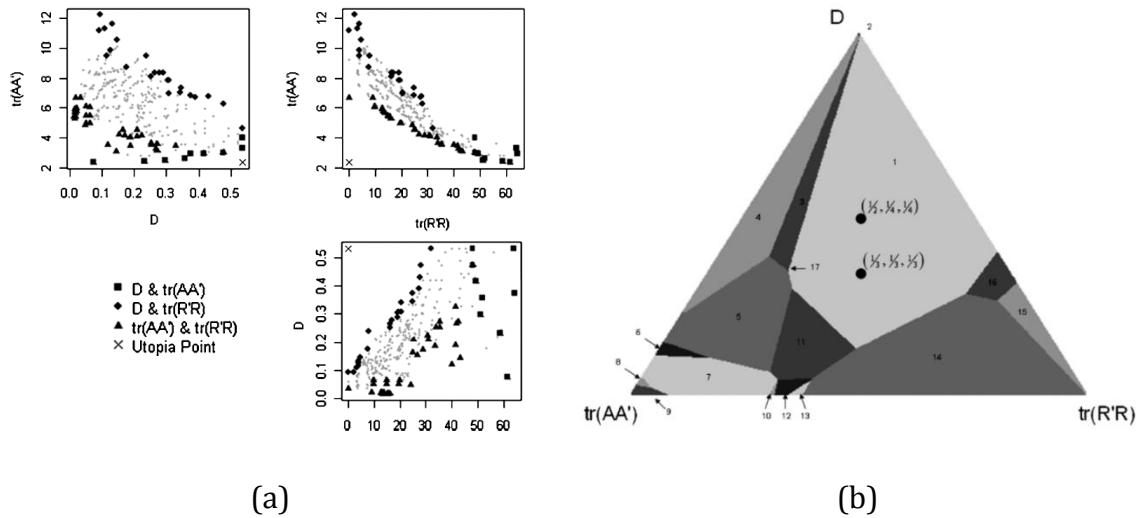


Figure 4.3 Pareto front results depicted in Lu et al. (2011)

### 4.3.2 Considering the Amount of Replication when Choosing a Design

Though the scatter and mixture plots provide a means to visualize the spread of criterion values for each of the designs on the front, also providing the number of  $df_{pe}$  for each of the competing designs is of practical value. As demonstrated in Chapter 1, the performance of an overall screening methodology also depends upon the ability of an analysis procedure to detect the truly active effects. Further, designs with many replicate points can severely impede the analysis procedure's performance if a sufficient number of unique points are not included in the design. Therefore, the choice of design and whether or not it should contain replicate design points should be influenced by the ability of the analysis procedure to properly make use of the resulting model-independent error estimate as well as its ability to effectively employ the capability of the unique points to estimate and discriminate competing models.

We illustrate this point in Figure 4.4 by re-producing the Lu et al. (2011) plots to include information concerning the number of replicated points for each design. From inspection of these new plots, we can see that designs with higher  $D_{eff}$  and lower  $tr(\mathbf{AA}')$  values have no replicated design points and thus do not permit a pairing with analysis procedures that make use of a model-independent error estimate. We can also see that strong preference to minimize  $tr(\mathbf{RR}')$  results in adopting a design with four replicate runs, which would be beneficial if planning to make use of the  $df_{pe}$  for analysis purposes; however, if this is not the intention, then using resources to carry out four replicated experimental runs should be otherwise justified or reasons for current preferences in criteria re-examined.

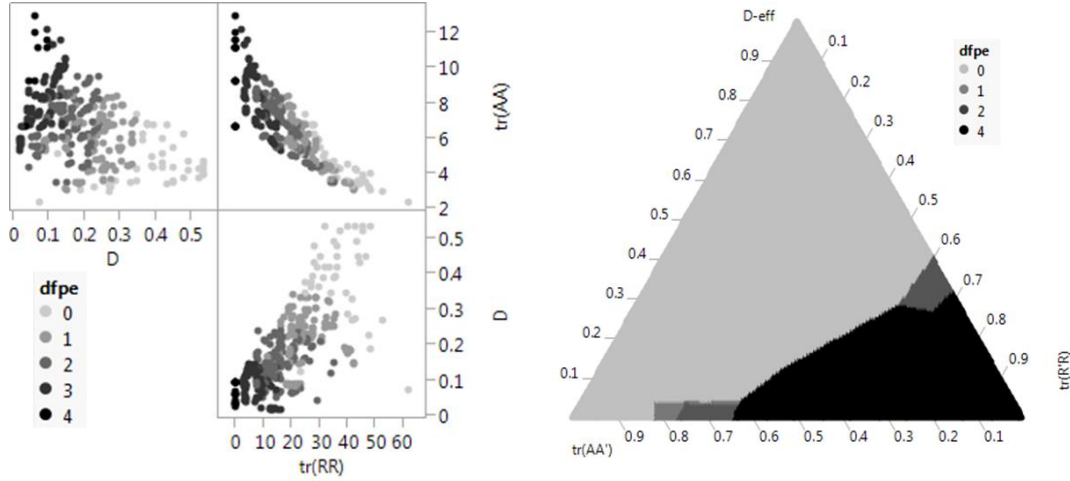


Figure 4.4 Reproduced results from Lu et al. (2011) colored by  $df_{pe}$

For the sequel of this chapter, we examine an implementation of the Pareto front approach to reduce the number of competing designs to a manageable size when screening for eight factors and their interactions using a total run size of  $n = 16$ . Two approaches are considered depending upon the intention of incorporating a model-dependent or independent error estimate into the analysis procedure. For the latter, the Bayesian  $DP$ -optimality criterion is incorporated into the PAPE algorithm along with  $tr(\mathbf{AA}')$  and  $tr(\mathbf{R}'\mathbf{R})$ . The combination of these criteria are meant to influence the creation of competing screening designs that complement an analysis strategy requiring an estimate of pure error while also helping to reduce the impact from model misspecification. We also implement the use of mixture plots to further reduce the set of Pareto front designs and provide rationale for selecting a final set of designs. We also explore the possibility of considering other designs not represented in the mixture plots by investigating their estimation capacity.

### 4.3.3 PAPE BD Screening Designs

The PAPE algorithm was implemented using  $tr(\mathbf{A}\mathbf{A}')$  and  $tr(\mathbf{R}'\mathbf{R})$  in combination with the Bayesian  $D$ -optimality criterion and the Bayesian  $DP$ -optimality criterion to form 681 unique PAPE  $BD$  and 487 unique PAPE  $BDP$  designs, respectively. Scatterplot matrices depicting the Pareto fronts formed by these two sets of results are given in Figure 4.5. When using either the Bayesian  $D$ -criterion or the Bayesian  $DP$ -criterion in combination with the  $tr(\mathbf{A}\mathbf{A}')$  and  $tr(\mathbf{R}'\mathbf{R})$  criteria, designs are produced which contain the highest amount of replication when  $tr(\mathbf{R}'\mathbf{R})$  is minimized, while the amount of replication is lowest when  $tr(\mathbf{A}\mathbf{A}')$  is minimized or either of the  $D$ -criterion are maximized. There also appears to be a wider selection of partially replicated designs with varying amounts of replication when compared to using the single-criterion approach as illustrated in Chapter 3.

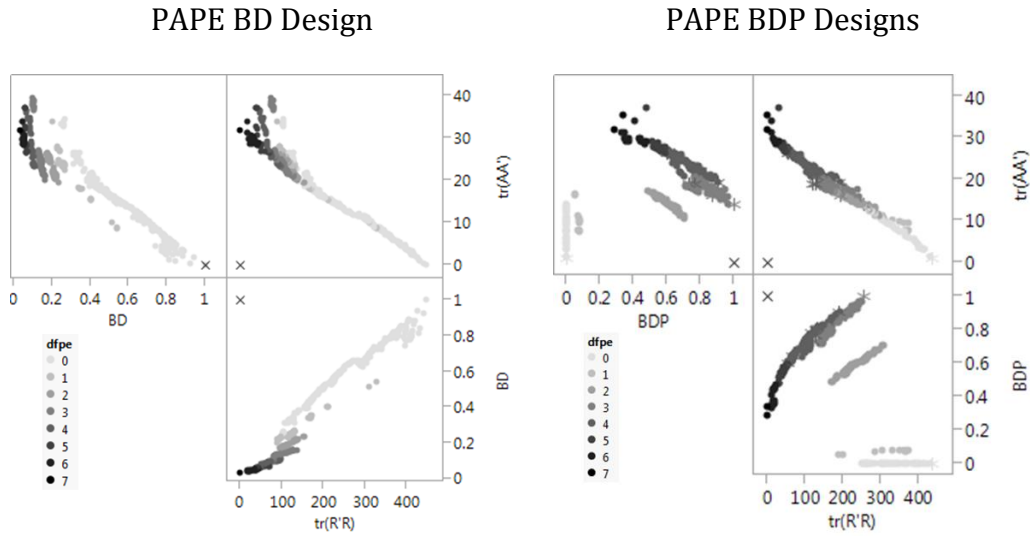
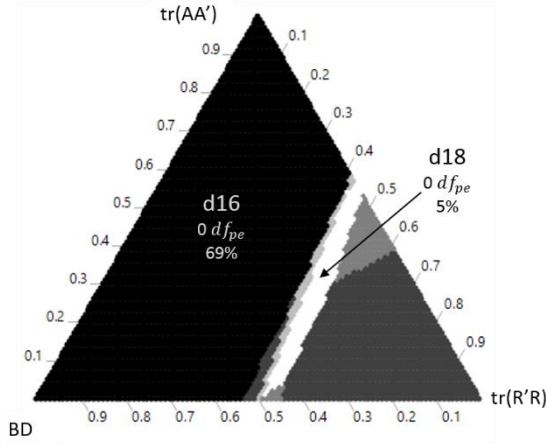


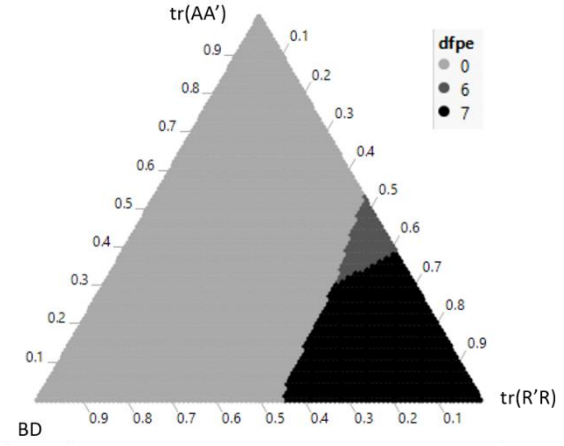
Figure 4.5 Scatterplot matrices of criterion values for PAPE designs

Mixture plots illustrating the reduced set of eight PAPE *BD* designs using the weighted sum score method of Lu et al. (2011) described in section 4.3.1 are shown in Figure 4.6 (a) colored by design and in Figure 4.6 (b) colored by number of available  $df_{pe}$ . Of the eight designs, six are un-replicated, one has 6  $df_{pe}$ , and one has 7  $df_{pe}$  making up 76%, 21%, and 3% of the total area of the mixture plot respectively. Although implementing the mixture plot method of Lu et al. (2011) reduced the number of competing Pareto front designs to roughly 1% of the total 681 initially created designs, the resulting choices for partially replicated designs contain a large number of replicate runs considering the total run size 20 runs.

The limitation imposed on the analysis procedure when not allowing for a sufficient number of unique points in the design was illustrated in Chapter 3, where the *DP*-optimal 7-factor design with 12 replicates heavily overfit the number of main effects and failed to detect most if not all the active two-factor interactions. For the work presented in this chapter, similar results were observed in initial simulations for the two 8-factor designs in question and are not presented since the focus of this chapter concerns implementing a multi-criterion approach. Therefore, we retain for later comparison only the two un-replicated PAPE *BD* designs since these designs were found to provide a practical comparison to alternatives used in this chapter when using an analysis procedures with a model-dependent error. These two designs correspond to those covering the largest amount of mixture plot area and can therefore be considered design choices more robust to uncertainty in criterion preference (Lu et al., 2011). These two designs are named d16 and d18, which are labeled in the Figure 4.6 (a).



(a) Colored by Design ID

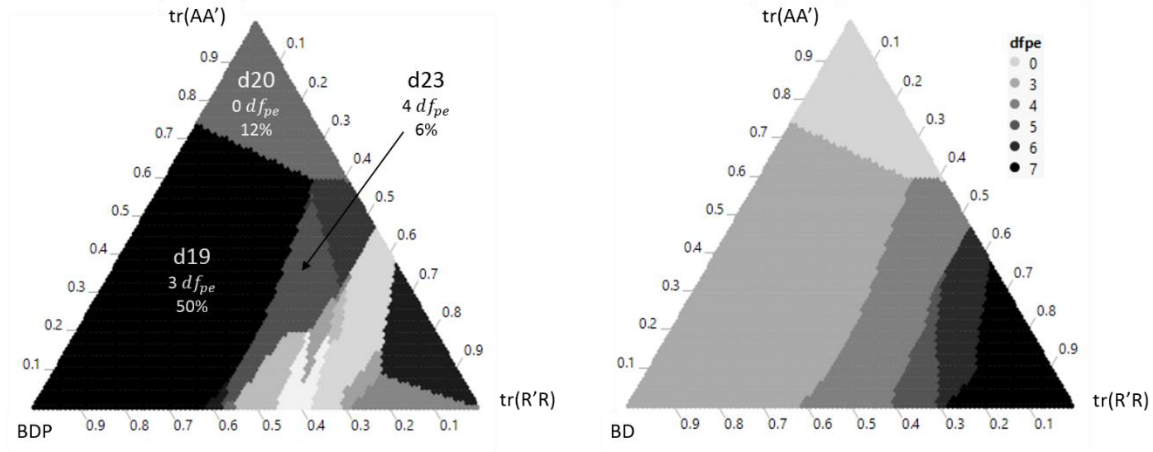


(b) Colored by  $df_{pe}$

Figure 4.6 Mixture plots for PAPE BD designs

#### 4.3.4 PAPE BDP Screening Designs

Extending the PAPE algorithm to incorporate the Bayesian  $DP$ -optimality criterion, we created 487 unique PAPE BDP designs which were reduced to a set of 19 designs using the weighted sum score method of Lu et al. (2011) described in section 4.3.1 and are presented in Figure 4.7 (a) as a mixture plot. In addition to having more design choices, we can also see in Figure 4.7 (b) that the PAPE algorithm paired with the Bayesian  $DP$ -criterion allows for a wider variety of possible  $df_{pe}$ . Of the 19 designs, only one design is un-replicated, two designs have 3  $df_{pe}$ , eight have 4  $df_{pe}$ , four have 5  $df_{pe}$ , two have 6  $df_{pe}$ , and two have 7  $df_{pe}$ , making up 12%, 50%, 18%, 3%, 7%, and 10% of the total area respectively. We retain for further comparison, the most  $BDP$ -efficient design (d19, 3  $df_{pe}$ ), the design with the lowest impact from  $tr(AA')$  (d20, 0  $df_{pe}$ ), and a design which incorporates a balance of all three criterion (d23, 4  $df_{pe}$ ). These designs are labeled in Figure 4.7 (a).



(a) Colored by Design

(b) Colored by  $df_{pe}$

Figure 4.7 Mixture plots for PAPE BDP designs

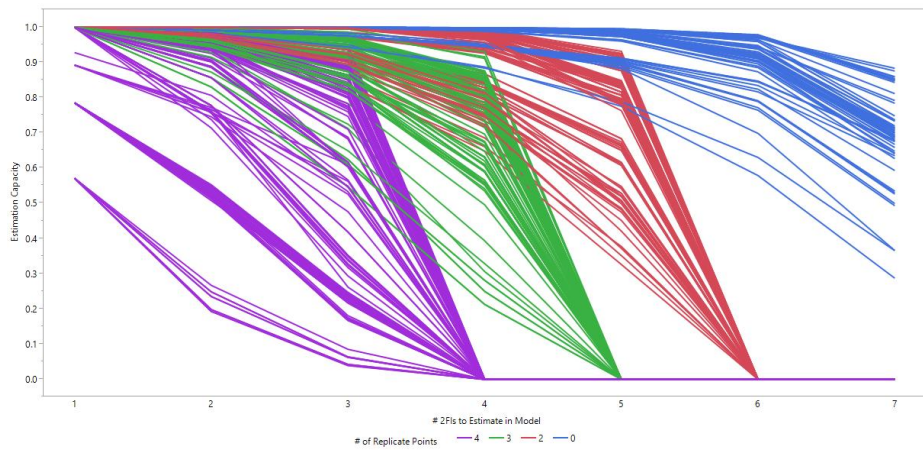
A small investigation was conducted looking at the estimation capacities (EC) of a subset of PAPE BDP designs found on the Pareto front. Estimation capacities (EC) for 336 PAPE BDP design containing two, three, or four replicates are depicted in Figure 4.8. Variation in EC can be seen for groups of designs containing the same number of replicate points, and the variation tends to increase as the number of replicates increase (or unique points decrease). Even designs with only four of the 16 runs replicated, there are large differences in EC. While the non-replicated designs allow for estimation of up to seven two-factor interaction effects in addition to all eight main effects, the partially replicated designs are each negatively impacted by the decreasing availability of  $df$  for estimating model terms. EC's for the eleven PAPE BDP designs illustrated in the mixture plot along with the two MRFD's previously introduced are given in Table 4.2. The three designs chosen from the mixture plot (d19, d20, and d23) do not necessarily provide the highest EC's. Further, comparison of the best PAPE BDP designs in terms of EC from Table 4.2 to those illustrated in Figure 4.8 indicates that better screening capabilities may be attainable from adopting



designs not provided in the mixture plots. To investigate this possibility we include an additional design, d49, as part of the collection of designs to be compared. This design was shown to have the highest EC sequence for designs containing four replicate points and its ability to estimate models containing one to three two-factor interactions was similar to those of the partially replicated MRFD suggested in Dasgupta et al. (2010).

*Table 4.2 Estimation capacities for 8-factor PAPE BD, PAPE BDP, and MRFD's*

Design	$df_{pe}$	$EC_1$	$EC_2$	$EC_3$	$EC_4$	$EC_5$	$EC_6$	$EC_7$
420 (20)	0	1	0.9921	0.9762	0.9473	0.8911	0.7719	0.4996
91	3	1	0.9603	0.8648	0.6314	0	0	0
15 (19)	3	1	0.8889	0.6252	0.2501	0	0	0
122	4	1	0.9550	0.8318	0	0	0	0
324	4	1	0.9339	0.6832	0	0	0	0
95	4	1	0.9048	0.6166	0	0	0	0
18	4	1	0.8889	0.6056	0	0	0	0
462 (23)	4	1	0.7778	0.4188	0	0	0	0
466	4	1	0.7778	0.4188	0	0	0	0
280	4	0.7857	0.5450	0.2387	0	0	0	0
351	4	0.5714	0.1984	0.0427	0	0	0	0
MRFD	4	1	0.9921	0.9228	0	0	0	0
MRFD	0	1	1	1	0.9993	0.9959	.9816	0.8999



*Figure 4.8 Estimation capacities for PAPE BDP designs*

### 4.3.5 Comparing BDP and PAPE BDP Designs

To investigate the claim of reducing the impact of model-misspecification when using the multiple-criterion approach, 200 designs created using solely the Bayesian  $DP$ -optimality criterion were compared in terms of  $tr(\mathbf{AA}')$  to 281 PAPE BDP designs having two to five replicate design points. Boxplots of the results are provided in Figure 4.9. Having been created with differently weighted preferences for criteria, the PAPE designs have different amounts of replication, whereas the BDP designs using a single criterion all have 4 replicates. Comparing these two sets of designs with 4 replicates indicates that there is not a noticeable impact on lowering  $tr(\mathbf{AA}')$ . PAPE BDP designs with two or three replicates show a reduction in  $tr(\mathbf{AA}')$  values, implying further protection against model-misspecification; however, further research is needed to determine how much of this reduction is due to the specific set of replicate points determined by the multi-criterion approach and how much of the reduction is simply due to the design now containing more unique design points.

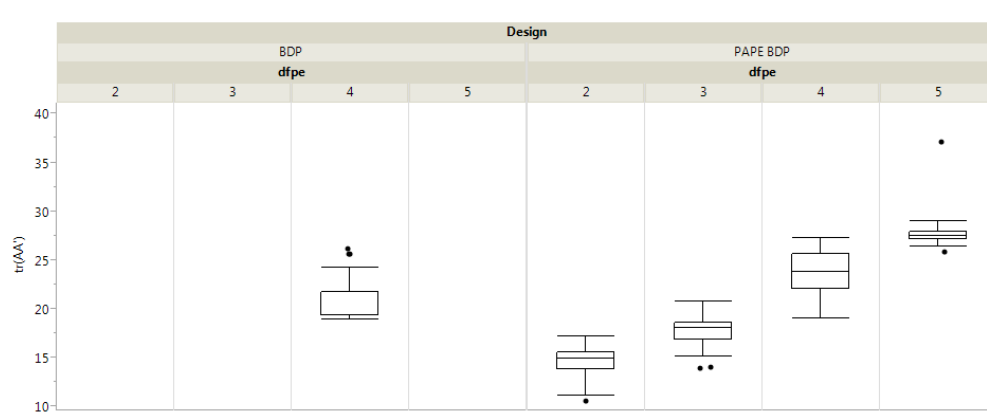


Figure 4.9 Boxplots of  $tr(\mathbf{AA}')$  for BDP and PAPE BDP Designs

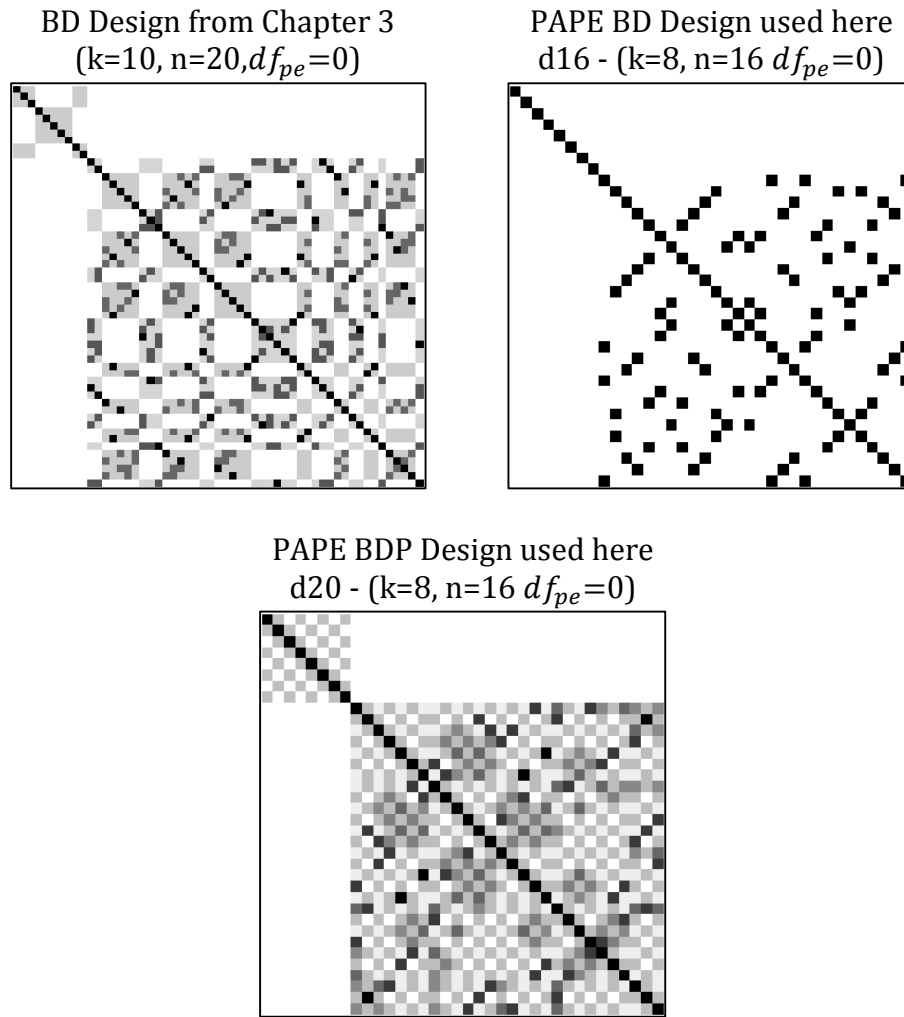
## 4.4 Simulation Results

Coverage probabilities and FDR's comparing the performance of the six PAPE designs to the MRFD's are provided in Figures 4.11 and 4.12 for when two and four of the eight main effects are active, respectively. The first four columns correspond to un-replicated designs being paired with either the Dantzig selector or forward selection. The last four columns correspond to partially-replicated designs being paired with either analysis procedure when using a model-dependent and model-independent error estimate. The mean number of main effect and two-factor interaction terms used to form the final model are provided in Figure 4.13. Power and type I error rates for main effects and two-factor interactions are given in Figure 4.14 and Figure 4.15 provides FDR's for the main effects and two-factor interactions.

When considering the four un-replicated designs - the PAPE *BD* designs d16 and d18; the PAPE BDP design d20; and the MRFD of Li and Natchsheim (2000).

- In terms of overall coverage (Figures 4.11 and 4.12, row 1), all designs appear to perform poorly.
- When paired with the Dantzig selector, d16 is the only PAPE design that consistently minimizes FDR's for dominant effects as the number of active interactions increases. This can be attributed to the Dantzig selector not detecting any two-factor interactions as active (Figure 4.13, row 2). Similar results were found in Chapter 3 when the procedure was paired with a 10-factor Bayesian *D*-optimal design (Figure 3.22, row 2). Correlation maps of the 10-factor Bayesian *D*-optimal and 8-factor PAPE Bayesian *D*-optimal designs are shown in the top two plots in Figure 4.10. Both

designs contain a structure which allows for main effects and two-factor interactions to be estimated independently from each other. As shown in the lower plot of Figure 4.10, the PAPE BDP design, d20, also used in this simulation study contains the same structure; however, d20 combined with the Dantzig selector does not exhibit the same behavior as the other designs and allows for the detection of some two-factor interactions (Figure 4.12, row 2 and Figure 4.13, rows 3 and 4).



*Figure 4.10 Correlation maps of BD designs with  $tr(\mathbf{A}\mathbf{A}')$  minimized*

- In terms of coverage for dominant and moderate effects (rows 2 and 3), the unreplicated PAPE BDP and MRFD appear to provide better coverage when paired with the Dantzig selector, and FDR's for dominant effects are also kept to minimal levels when combining the Dantzig selector with the MRDF (row 6). It also appears that the MRFD combination does not suffer as much from a problem of under-fitting the number of two-factor interactions (Figure 4.13, row 2), although FDR's for two-factor interactions indicate that up to a half are incorrectly identified (Figure 4.15, row 2)

Considering the four partially-replicated designs - the PAPE *BD* design d19 with 3  $df_{pe}$ ; the PAPE *BDP* designs d23 & d49, each with 4  $df_{pe}$ ; and the augmented MRFD of Dasgupta et al. (2010) with 4  $df_{pe}$ .

- Forward selection does not appear to benefit from the use of a model-independent error estimate since the slightly higher coverages are at the expense of higher FDR's for all effect sizes; however, in a screening situation, a practitioner may allow for type I errors in order to detect the active effects. Forward selection performs similarly when paired with any of the partially replicated designs including the MRFD with 4  $df_{pe}$  and d49, which was not contained in the mixture plot but included in the study since has higher estimation capacity than other design in the mixture plot.
- Whether or not an independent error estimate is used, the performances of the Dantzig selector appear indistinguishable for any given design.

The first and second bullets imply that model-independent analysis procedures may not be a practical choice when screening eight factors and their interactions, and further suggests that partial replication of screening designs should perhaps be limited to cases where they

can provide a complementary benefit to paired analysis procedures. However, additional modification of either analysis procedure or the adoption of another more suitable for being paired with partially replicated designs may suggest otherwise.

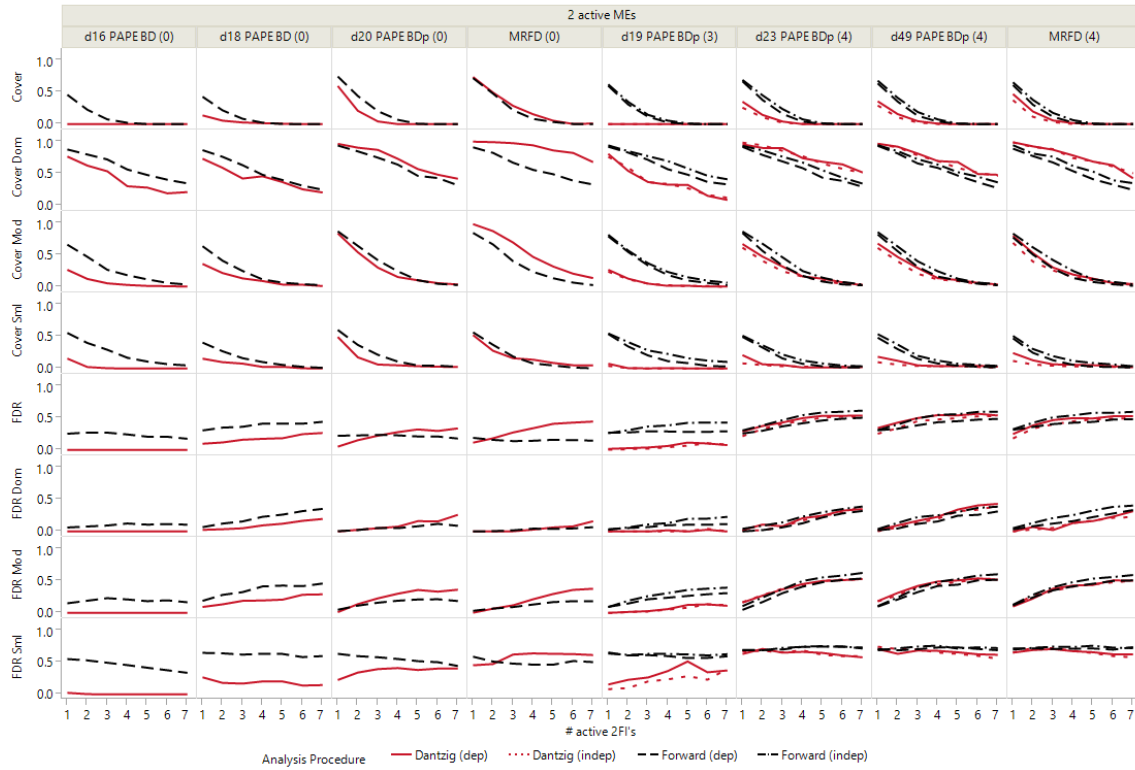


Figure 4.11 Cover and FDR's when 2/8 main effects are active ( $n=16$ )

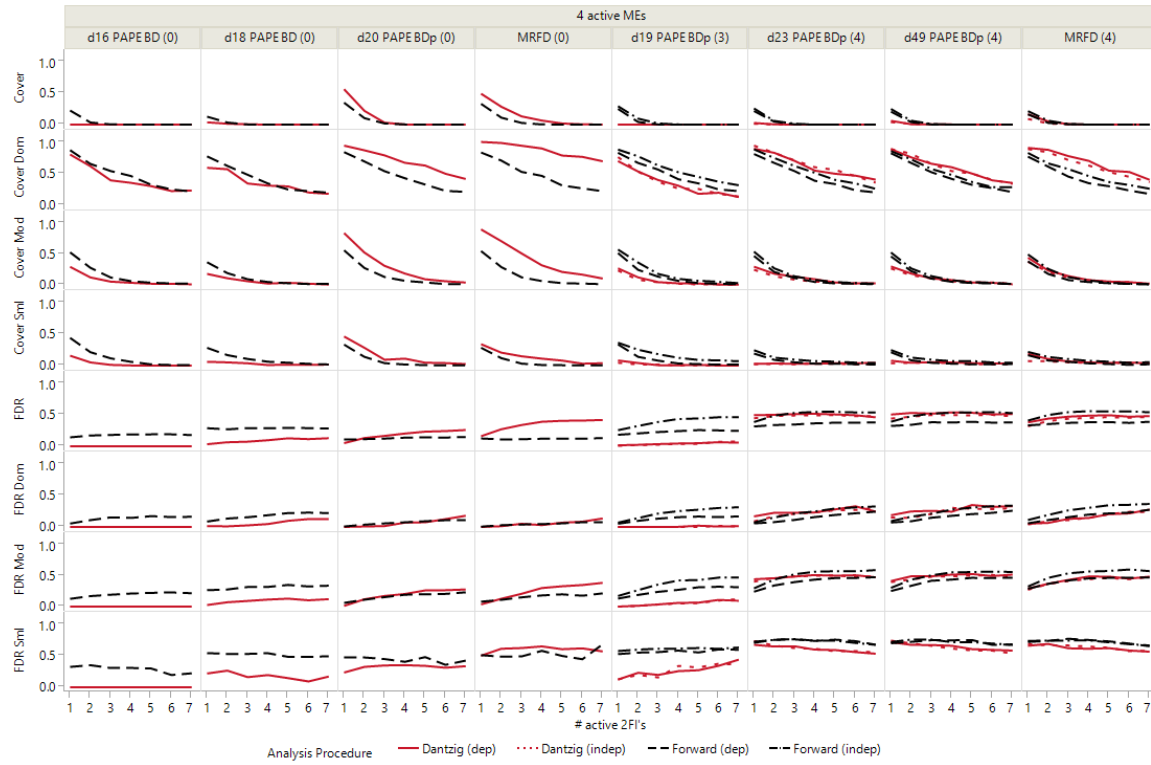


Figure 4.12 Cover and FDR's when 4/8 main effects are active ( $n=16$ )

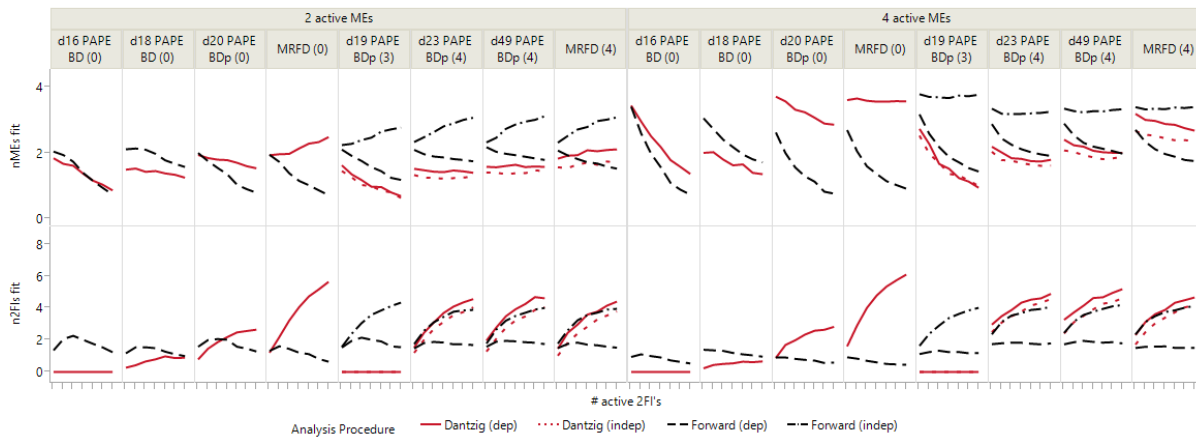


Figure 4.13 Number of terms in final model ( $k=8$ ,  $n=16$ )

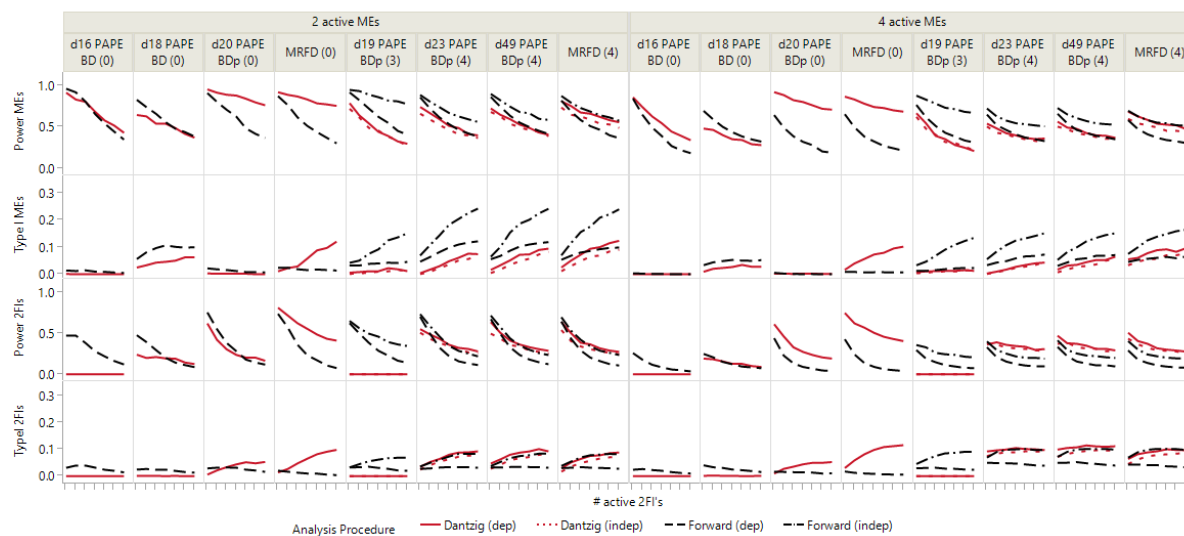
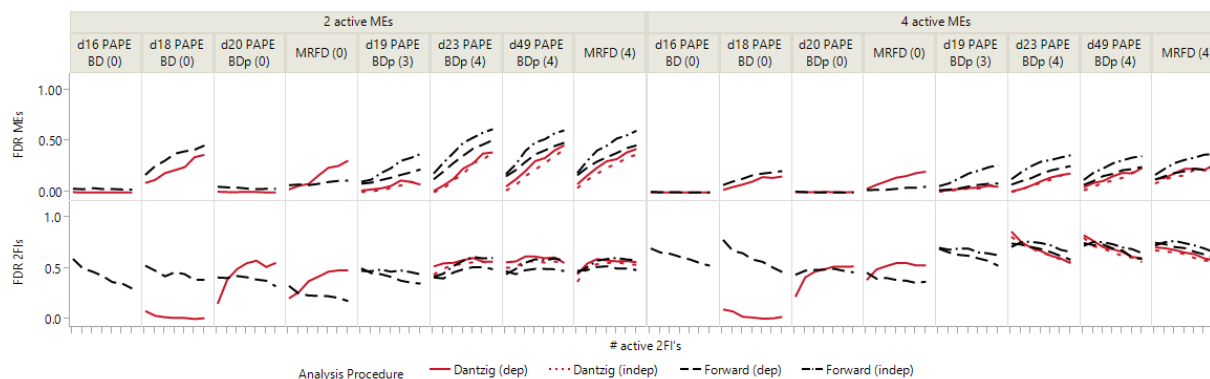


Figure 4.14 Power and type I error rates for effects ( $k=8, n=16$ )



4.15 FDR's for main effects and two-factor interactions ( $k=8, n=16$ )



## 4.5 Comments and Discussion

We have illustrated how incorporating the Bayesian  $DP$ -criterion into the PAPE algorithm supports the flexibility of an optimization approach to create partially replicated screening designs when more than one criterion is to be used. An analysis of the Pareto front designs when using the Bayesian  $DP$ -optimality criterion did not show a clear impact from using multiple criteria simultaneously to create designs. Correlations between effects and  $tr(\mathbf{AA}')$  values for 200 BDP designs and 139 PAPE BDP designs were found to be similar when they contain the same number of replicate design points. There were noticeable decreases in  $tr(\mathbf{AA}')$  values for PAPE BDP designs with fewer numbers of replicate points; however, the impact on improving protection against model-misspecification is not clear.

Simulation results show the model robust factorial design of Li and Natchsheim (2000) outperforms the proposed un-replicated PAPE designs. Further, designs based upon the Bayesian  $DP$ -optimality criterion are shown to perform similarly to the augmented MRFD's found in Dasgupta et al. (2010). It is interesting to find that the un-replicated PAPE design, d20, produced by the incorporation of the Bayesian  $DP$ -criterion showed better performance both in terms of coverage and FDR's than either of the PAPE designs based upon the Bayesian  $D$ -criterion (d16, d18). That is, it was surprising to find that the implementation of a criterion meant to produce designs with replication actually resulted in a design that performed better than either of the non-replicated designs that are based upon a criterion that limits replication of points. The structure of one of the PAPE BD designs, d16, was found to have a similar structure to that of the 10-factor BD design used in Chapter 3; however, the PAPE BDP design, d20, used in this chapter also contains a similar structure. The EC's of these designs reveal that the sequences of EC's for d16 and d18 are far lower

than d20, beginning with the consideration for model sizes up to two-factor interactions. Further research is planned to investigate this result.

Inspection of estimation capacity for BDP designs on the Pareto front reveals that EC can vary widely for BDP designs having the same number of replicate points, and that for a given run size the variation in EC of partially replicated BDP designs increases as the number of unique design points decreases. A BDP design with four replicate points was chosen from the Pareto front that was not contained in the mixture plot, and compared in the simulation. Although this design shows higher EC than other designs in the mixture plot, results did not show any improved performance when using this design.

Proposed improvements in the use of mixture plots by incorporating the amount of replication within each design were shown to provide a practical benefit to the design selection process. The influence of  $tr(\mathbf{R}'\mathbf{R})$  was discovered to directly contribute to increased amounts of replication in designs. Though the amount of replication is extreme for screening purposes, further research into the combined influence of  $tr(\mathbf{R}'\mathbf{R})$  and with other criteria to create designs more robust to model-misspecification is needed. Further investigation of different criteria combinations and different methods to control the PAPE algorithm's searching mechanism are also two areas of possible exploration.

## Chapter 5

# Considerations for Follow-Up Experimentation

### 5.1 Introduction

The use of follow-up experimentation to improve the ability to discriminate between models can be a practical consideration since initial run sizes in some situations permit estimation of all models of interest, some of which may only include some interactions and quadratic effects if a priori knowledge of the system is reliable. However, in many cases there is limited or no knowledge of the system and follow-up experimentation is used in order to improve upon the knowledge gained from previous experiments. Ruggoo and Vandebroek (2004) show that a single follow-up experiment can have major impact on improving system understanding even if initial knowledge is poor. Investigating *D*-optimal follow-up procedures, Silvestrini (2013) also shows that assuming a simpler form of an initial model can lead to better results than starting with a more complex model assumption. However, in some screening situations the results from an initial experiment may not provide reliable information even when a simple model form is assumed. For supersaturated designs, having a small run size relative to the number of factors being investigated restricts interest to estimating main-effect only models where even a subset of these can likely result in a large number of models explaining the data well (Edwards and Mee, 2011, Abraham et al., 1999, Sunburg, 2008). However, for supersaturated model cases, where estimation of main-effects and some two-factor interactions is possible, follow-up experimentation can be a powerful

tool if properly implemented. In this chapter we will investigate the possible benefits and limitations of using a model-independent error estimate when a two-stage sequential experiment is considered. We compare performance of different sequential screening approaches in order to determine if replication of design points is best suited for the initial phase or the follow-up of the sequential process. Bayesian *DP*-optimal augmentation is presented and incorporated into the method proposed by Gutman et al. (2014), where considerations for secondary terms are used to help identify active two-factor interactions not detected in the initial screening.

## 5.2 Optimal Determination of Follow-Up Runs

### 5.2.1 *D*-Optimal Follow-Up

The objective of follow-up experimentation is to take what was learned from the previous experiment or series of experiments and use this knowledge to form another set of runs in order to further improve understanding of the system being investigated. Using *D*-optimality is one way of determining a set of follow-up experimental runs. Methods to implementing this approach can be found in Goos and Jones (2011) and Atkinson et al. (2007). Suppose after an initial experiment, a newly specified model is to be estimated using the addition of follow-up experimental runs. Let  $\mathbf{X}_i$  be the  $n_i \times q$  model matrix corresponding to the first  $n_i$  runs of the initial experiment, and let  $\mathbf{X}_f$  be the  $n_f \times q$  model matrix corresponding to the follow-up design with  $n_f$  runs, where  $n = n_i + n_f$  and

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_i \\ \mathbf{X}_f \end{bmatrix} \quad (5.1)$$

is the final  $n \times q$  model matrix for the *a priori* specified model corresponding to the complete design. Implementing  $D$ -optimality, we then seek to maximize the determinant of the complete information matrix,

$$\begin{aligned}\mathbf{X}'\mathbf{X} &= \begin{bmatrix} \mathbf{X}_i \\ \mathbf{X}_f \end{bmatrix}' \begin{bmatrix} \mathbf{X}_i \\ \mathbf{X}_f \end{bmatrix} \\ &= [\mathbf{X}_i' \mathbf{X}_f'] \begin{bmatrix} \mathbf{X}_i \\ \mathbf{X}_f \end{bmatrix} \\ &= \mathbf{X}_i' \mathbf{X}_i + \mathbf{X}_f' \mathbf{X}_f .\end{aligned}$$

Therefore, a  $D$ -optimal follow-up design maximizes

$$|\mathbf{X}'\mathbf{X}| = |\mathbf{X}_i' \mathbf{X}_i + \mathbf{X}_f' \mathbf{X}_f| \quad (5.2)$$

over all possible matrices  $\mathbf{X}_f$  since the initial experiment has already been conducted. That is, since the rows of  $\mathbf{X}_i$  are known and are therefore fixed, a design construction algorithm only needs to consider possible exchanges corresponding to the design points in  $\mathbf{X}_f$ .

### 5.2.2 Bayesian $DP$ -Optimal Follow-Up

Recall from Chapter 3 that a drawback to the  $D$ -optimality approach is the required assumption that the assumed model is known and that to overcome the dependency on an assumed model, we introduced the simple Bayesian modification of DuMouchel and Jones (1994) which involves the defining of *primary* and *potential* terms of interest. Incorporating such an approach, Ruggoo and Vandebroek (2004) extend the work of Neff (1996) to develop two-stage Bayesian  $D$ - $D$  optimal design where improved model knowledge from the initial experiment (first-stage) is used to generate a follow-up experiment (second-stage). This is done by updating the prior covariance matrix with new information gathered from the initial experiment. Recall from Chapter 3 that  $\mathbf{R} = \mathbf{K}/\tau^2$  is a prior covariance matrix and that the

prior distribution of the unknown model parameters is assumed to be  $\boldsymbol{\beta}|\sigma^2 \sim N(\boldsymbol{\beta}_0, \sigma^2 \mathbf{R}^{-1})$ .

Let  $\mathbf{y}_i$  and  $\mathbf{y}_f$  be the initial and follow-up observation vectors, respectively. Then

assuming  $\mathbf{y}_f|(\boldsymbol{\beta}, \sigma^2) \sim N(\mathbf{X}_f \boldsymbol{\beta}, \sigma^2 \mathbf{I}_{n_f})$  as the conditional distribution of  $\mathbf{y}_f$  given  $\boldsymbol{\beta}$  and  $\sigma^2$ ,

Ruggoo and Vandebroek (2004) show that

$$\boldsymbol{\beta}|\mathbf{y} \sim N\left[\hat{\boldsymbol{\beta}}, \sigma^2(\mathbf{X}_i' \mathbf{X}_i + \mathbf{X}_f' \mathbf{X}_f + \mathbf{R})^{-1}\right] \quad (5.3)$$

is the posterior distribution for  $\boldsymbol{\beta}$  given  $\mathbf{y} = \begin{pmatrix} \mathbf{y}_i \\ \mathbf{y}_f \end{pmatrix}$ , where  $\hat{\boldsymbol{\beta}} = (\mathbf{X}_i' \mathbf{X}_i + \mathbf{X}_f' \mathbf{X}_f + \mathbf{R})^{-1}(\mathbf{X}_i' \mathbf{y}_i +$

$\mathbf{X}_f' \mathbf{y}_f + \mathbf{R} \boldsymbol{\beta}_0)$ . This methodology was later adapted to Bayesian  $D$ -optimal augmentation of

supersaturated designs by Gutman et al. (2014), who gave further consideration to

specifying *secondary* terms, in addition to *primary* and *potential* terms. Because they focus

on supersaturated designs, Gutman et al. do not include higher order terms such as

interaction or quadratic effects. That is, they specify each of the  $k$  factor and intercept model

terms as being one of the  $p_1$  primary,  $p_2$  secondary, or  $p_3$  potential terms, where  $p = p_1 +$

$p_2 + p_3$ . Guidelines suggested in Gutman et al. (2014) when making these specifications are,

1. Always specify the intercept as primary
2. Classify all terms as potential if there is high uncertainty in making the classifications
3. If analysis suggests  $p < n_i + n_f$  terms as active, then specify these terms as primary
4. If the number of model terms of interest is  $p > n_i + n_f$ , then specify all  $p$  terms as secondary terms
5. Any terms with little or no supportive evidence as being active are specified as potential.

By updating the specification of model terms in  $\mathbf{R}$  as primary, potential, or secondary using knowledge gathered from the initial experiment, Gutman et al. (2014) determine a set of Bayesian  $D$ -optimal follow-up runs by choosing  $\mathbf{X}_f$  to maximize the determinant

$$|\mathbf{X}'_i \mathbf{X}_i + \mathbf{X}'_f \mathbf{X}_f + \mathbf{R}|, \quad (5.4)$$

where

$$\mathbf{R} = \frac{\mathbf{J}}{\gamma^2} + \frac{\mathbf{K}}{\tau^2}$$

$$= \frac{1}{\gamma^2} \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & j_{1,1} & & & \\ 0 & & j_{2,2} & & \\ \vdots & & & \ddots & \\ 0 & \mathbf{0} & & & j_{k,k} \end{bmatrix} + \frac{1}{\tau^2} \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & k_{1,1} & & & \\ 0 & & k_{2,2} & & \\ \vdots & & & \ddots & \\ 0 & \mathbf{0} & & & k_{k,k} \end{bmatrix}.$$

Here we can see that the prior covariance matrix has been re-structured as the sum of two individually scaled indicator matrices  $\mathbf{J}$  and  $\mathbf{K}$  to allow for the specification of secondary and potential terms respectively. Scaling factors,  $\gamma$  and  $\tau$ , are chosen to represent beliefs in certain factors, where larger values correspond to stronger beliefs that certain factors are active. Gutman et al. use  $\gamma^2 = 5$  as suggested in Jones et al. (2008) and  $\tau^2 = 100$ , since these settings were found to be robust to model misspecification. Zero vectors make up the first row and column of each  $\mathbf{J}$  and  $\mathbf{K}$  matrix since the intercept model term is always specified as primary, and therefore would not be indicated by a setting of 1 in either matrix. Specifications of the  $x_i$  model terms for  $i = \{1, 2, \dots, k\}$  as  $p_2$  secondary terms or  $p_3$  potential terms are made by setting the corresponding  $j_{i,i} = 1$  or  $k_{i,i} = 1$ , respectively. All remaining model terms not specified as either secondary or potential are specified as primary terms by setting  $j_{i,i} = 0$  and  $k_{i,i} = 0$ .

To allow for an optimal determination of partially replicated follow-up design points, we extend (5.4) to incorporate  $DP$ -optimality and propose that a set of Bayesian  $DP$ -optimal follow-up runs be determined by maximizing

$$|\mathbf{X}'_i \mathbf{X}_i + \mathbf{X}'_f \mathbf{X}_f + \mathbf{R}| / (F_{p,d:1-\alpha})^p \quad (5.5)$$

where  $p = (n_i + n_f) - 1$ ,  $d$  is pure error degrees of freedom, and  $F_{p,d:1-\alpha}$  is the  $(1 - \alpha)$ -quantile of the  $F$ -distribution. By incorporating (5.5) into the sequential process, an optimal number of replicate design points can now be determined for follow-up procedures. To investigate the potential benefit and limitation of using Bayesian  $DP$ -optimal augmentation for follow-up experimentation, we conducted two simulation studies focusing on the impact of using partially replicated follow-up runs where the use of a model-independent error may have a more positive impact on the overall screening methodology. Details concerning the two simulations are provided in the next section.

### 5.3 Simulation Settings for Sequential Experimentation

Both simulations make use of two MRFD's from Li and Nachtsheim (2000), where one is unreplicated and the other partially replicated. These designs were chosen in order to limit the amount of variability in results due to different design characteristics while still permitting adequate ability to detect active effects when paired with forward selection if using a model-dependent error estimate. The first study uses 8-factor designs and the second uses 7-factor designs, and both use  $n_i = 16$  initial runs with follow-up run sizes of  $n_f = \{4, 8\}$ . In both simulations, 2 or 6 main effects are simulated as active in addition to 1 to 7 two-factor



interactions. We carry out the following simulation in each of 2,000 iterations using forward selection:

6. Steps 1 to 5 as described in Chapter 2, section 5 are carried out to simulate the initial experiment. If the initial design is un-replicated, then a model-dependent error estimate is used, otherwise a model-independent error is used.
7. The results from step 6 are then used to update the prior covariance matrix,  $\mathbf{R}$ .
8. After incorporating the updated information in  $\mathbf{R}$ , a set of  $n_f$  follow-up runs are determined by one of the following two methods:
  - Bayesian  $DP$ -optimality to form a set of runs where replication is determined by the criterion.
  - A fixed set of replicate points is forced using Bayesian  $D$ -optimality to first form a set of  $n_u$  unique design points and then a set of  $n_f - n_u = n_d$  replicate points is determined using the Bayesian  $D$ -optimality criterion.
9. Letting  $\mathbf{X}_f$  be the follow-up matrix, and letting  $\mathbf{X}'_a$  be the matrix consisting of the  $m$  and  $t$  columns  $\mathbf{X}_f$  corresponding to the truly active effects, a single follow-up  $n_f \times 1$  response vector is formed as  $\mathbf{y}_f = \mathbf{X}'_a \boldsymbol{\beta} + \boldsymbol{\epsilon}$ , where  $\boldsymbol{\epsilon} \sim N(0,1)$ .
10. Analysis of the follow-up experiment is conducted.

For the first simulation, two 8-factor MRFD's from Li and Nachtsheim (2000) are used as initial screening designs. The first is a 16-run un-replicated MRFD, and the second is a 16-run partially replicated design from Dasgupta et al. (2010), created by augmenting a 12-run MRFD with four replicate runs using A-optimality. Correlation maps for these designs are illustrated in Figure 5.1 and designs provided in Table 5.1. The aliasing between main effects and two-factor interactions is noticeable lower in the un-replicated MRFD.

Because the Bayesian  $DP$ -criterion optimally determines the number of replicate design points, we compare the implementation of (5.5) to follow-up approaches where the number of replicate points is fixed. Beginning with the initial 16-run screening experiment, follow-up run sizes of  $n_f = \{4,8\}$  are used, forming a complete two-stage sequential experiment in either 20 or 24 total runs. In the cases where (5.5) is not implemented, a set

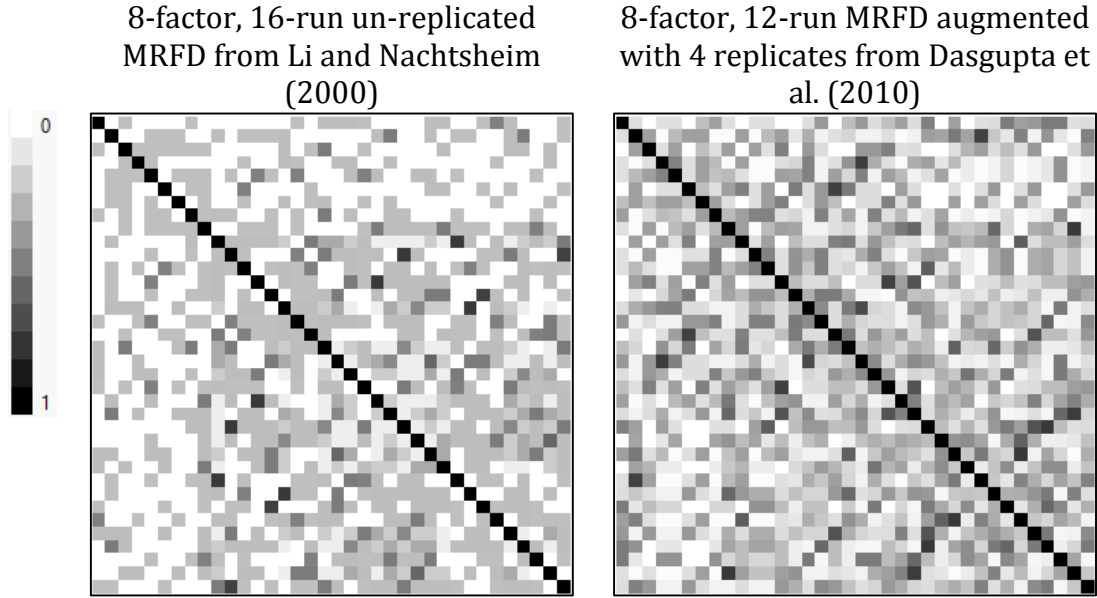


Figure 5.1 Correlation maps for 8-factor initial designs

Table 5.1 Un-replicated and partially replicated 8-factor MRFD's

Un-replicated MRFD							
$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$	$x_8$
-1	1	-1	1	1	1	1	-1
1	1	1	1	1	-1	-1	-1
1	1	1	1	-1	1	1	1
1	-1	1	-1	-1	-1	-1	-1
-1	-1	-1	-1	1	1	1	1
1	-1	-1	1	1	-1	1	1
-1	1	-1	1	-1	1	-1	1
-1	-1	-1	-1	1	-1	-1	-1
-1	-1	1	1	-1	-1	-1	1
1	-1	1	1	-1	1	-1	-1
-1	1	-1	-1	-1	-1	1	1
1	1	-1	-1	1	1	1	-1
1	1	1	-1	1	-1	1	-1
-1	-1	1	1	-1	-1	1	-1
1	-1	-1	-1	-1	1	-1	1
-1	1	1	-1	1	1	-1	1

Partially Replicated MRFD							
$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$	$x_8$
1	1	-1	-1	1	1	-1	1
1	-1	-1	1	-1	1	1	-1
1	1	1	1	1	-1	-1	-1
-1	1	1	-1	1	1	-1	-1
-1	-1	1	1	-1	-1	1	-1
1	-1	1	-1	1	-1	-1	1
-1	1	1	1	1	1	1	-1
-1	-1	-1	1	-1	1	1	1
-1	-1	-1	-1	-1	-1	-1	1
1	1	-1	1	-1	-1	-1	1
-1	-1	-1	-1	1	-1	1	-1
1	1	1	-1	-1	1	1	1
1	-1	-1	1	-1	1	1	-1
-1	-1	-1	-1	1	-1	1	-1
-1	-1	-1	-1	1	-1	1	-1

of partially replicated follow-up runs for a fixed number of replicate points is determined using a two-step process where the  $n_u$  non-replicated follow-up runs are first determined using (5.4) and the remaining  $n_f - n_u$  follow runs determined using a candidate-exchange algorithm with Bayesian  $D$ -optimality and an updated prior covariance matrix. For the cases where  $n_f = 4$ , we replicate two of the follow-up runs. For the cases where  $n_f = 8$ , we simulate two separate cases where either two or four of the follow-up runs are replicated.

In addition to varying the amount of replication in the follow-up design, we also incorporate two methods of updating the prior covariance matrix and values for the scaling factors. The first method specifies model terms identified as active from the initial experiment in addition to the intercept as primary terms and all other terms as potential and a scaling factor of  $\tau^2 = 1$  is used. For the second method, an adaptation of the guidelines in Gutman et al. (2014) in order to consider some two-factor interactions as secondary terms. With this method actively identified terms from the initial experiment in addition to the intercept are specified as primary terms, secondary terms are specified as being any two-factor interaction with at least one parent effect specified as primary and all other terms (possibly including some main effects) as potential. For this method, scaling factors  $\gamma^2 = 100$  and  $\tau^2 = 5$  as suggested in Gutman et al. (2014) are used. To illustrate the effect of this approach, correlation maps for the 8-factor 16-run MRFD in Li and Nachtsheim (2000) and a 20-run complete design from a 4-run Bayesian  $D$ -optimal augmentation are shown in Figure 5.2. The 16-run initial MRFD and the follow-up design using four runs are shown in Table 5.2.

*Table 5.2 Initial and follow-up designs for eight factors*

$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$	$x_8$	
-1	1	-1	1	1	1	1	-1	Initial 16-run MRFD
1	1	1	1	1	-1	-1	-1	
1	1	1	1	-1	1	1	1	
1	-1	1	-1	-1	-1	-1	-1	
-1	-1	-1	-1	1	1	1	1	
1	-1	-1	1	1	-1	1	1	
-1	1	-1	1	-1	1	-1	1	
-1	-1	-1	-1	1	-1	-1	-1	
-1	-1	1	1	-1	-1	-1	1	
1	-1	1	1	-1	1	-1	-1	
-1	1	-1	-1	-1	-1	1	1	
1	1	-1	-1	1	1	1	-1	
1	1	1	-1	1	-1	1	-1	
-1	-1	1	1	-1	-1	1	-1	
1	-1	-1	-1	-1	1	-1	1	
-1	1	1	-1	1	1	-1	1	
1	-1	1	-1	-1	-1	1	1	4-run follow-up design
-1	1	1	1	1	-1	1	1	
1	1	-1	-1	1	-1	-1	1	
-1	-1	1	-1	-1	1	1	-1	

The four follow-up runs were determined using the Bayesian  $D$ -optimality criterion and assuming the initial screening analysis indicated main effects  $x_1$  and  $x_2$  along with interaction  $x_1x_2$  as active and updating the prior covariance matrix using method 2. In the correlation map on the left in Figure 5.2, the two main effects,  $x_1$  and  $x_2$ , along with their interaction are shown to be uncorrelated. The correlation map on the right shows that augmenting the initial design with four follow-up runs retains the independence between the three primary effects, while some correlations among the remaining main effects have increased. In general, the correlations between the main effects and secondary terms (any interaction containing either  $x_1$  or  $x_2$ ) has remained low, while the correlations between the main effects and interactions not involving either  $x_1$  or  $x_2$  have increased.

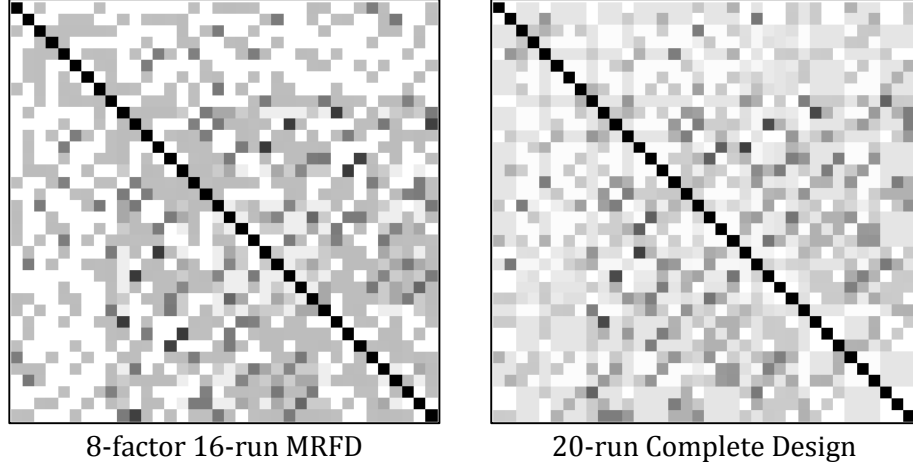


Figure 5.2 Correlation maps of initial design and complete design

## 5.4 Simulation Results

### 5.4.1 8-factor Designs using $n_i = 16$

Starting with an un-replicated or partially replicated 8-factor design did not have a noticeable impact on the simulation results once follow-up runs were added and analysis conducted. However, the choice of updating method was found to have a small influence when using either design. This is illustrated using results when the initially un-replicated designs is used. The top row of Figure 5.3 displays the average number of  $df_{pe}$  of the designs used in the simulation while the bottom row provides average RMSE. We can see in the initial experiment that forward selection overestimates RMSE since the initial design (same throughout the simulation during this first stage) is un-replicated. Further, as the number of active main effects and two-factor interactions increase, the overestimation becomes worse; however, since the follow-up experiments allow for replication of design points, a model-independent estimation of error is shown to properly represent the systematic error used in the simulation (i.e.,  $\epsilon \sim N(0,1)$ ). Focusing on the top two middle plots in Figure 5.3, we can

see when four follow-up runs are used, that incorporating the Bayesian *DP*-optimality criterion into either follow-up method results in all four follow-up runs being replicate design points. However, when eight additional runs are instead used to form a follow-up experiment, the criterion is influenced differently depending on which updating method is used. When using method 2, there are fewer replicates deemed necessary to estimate pure error and as the number of active two-factor interactions increases the number of replicates slightly increases as well. Implementing method 2 translates into four to five points being replicated rather than a consistent six replicates when using method 1. With fewer replicates, the resulting combined design has more unique points to help resolve aliasing and bias issues stemming from a small run size.

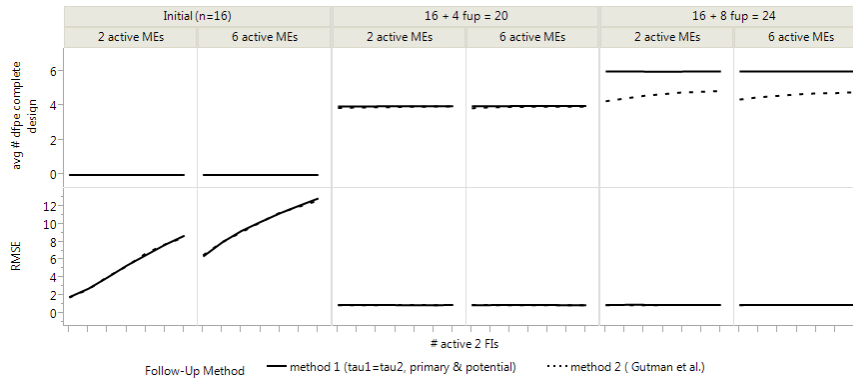
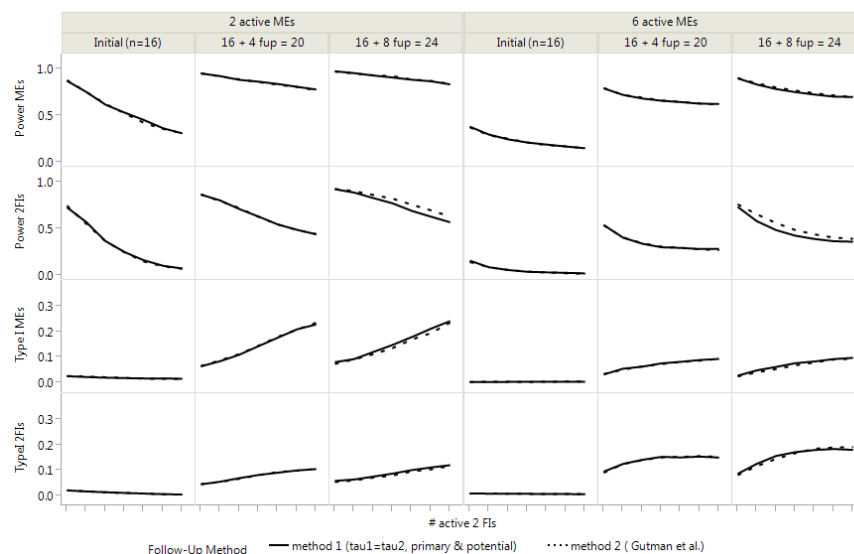


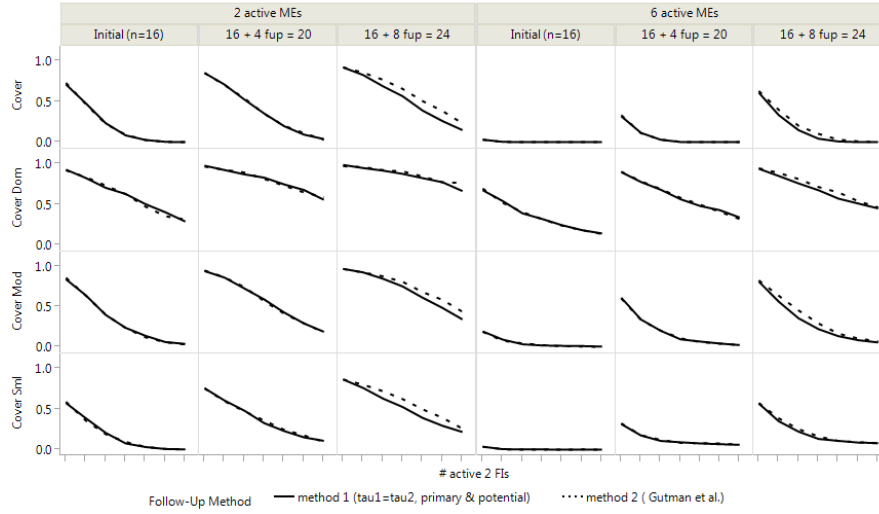
Figure 5.3 Average number of  $df_{pe}$  and RMSE for 8-factor case

To illustrate the impact of this result on follow-up performance, we provide simulation results for power and type I error rates for main effects and two-factor interactions in Figure 5.4 as well as coverage probabilities for the three effect sizes in Figure 5.5. The similar performances shown in any of the figures when adding four follow-up runs reflect the likely result of all four follow-up runs being replicate design points when using

either updating method. Results also indicate that power to detect main effects is not influenced by the choice of updating method although there is a slight decrease in type I error rates. The more noticeable result is the increased power to detect active two-factor interactions. The improvement also appears to take place depending upon the structure of the system model. That is, when two main effects are active there is an increase in power to detect two-factor interactions when there are a larger number active. Conversely, when there are six main effects active, there is an increase in power to detect active two-factor interactions only when there are few that are actually active. The increase in coverages corresponding to the eight-run follow-up shown in Figure 5.5 indicate that the detection of small to moderate sized two-factor interactions is contributing to the increased power shown in Figure 5.4 when two main effects are active, while increased detection of moderate to dominant sized two-factor interactions is contributing to the increased power when six main effects are active. This is due to active two-factor interactions more likely being assigned larger coefficient values since there are more main effects being simulated as active.



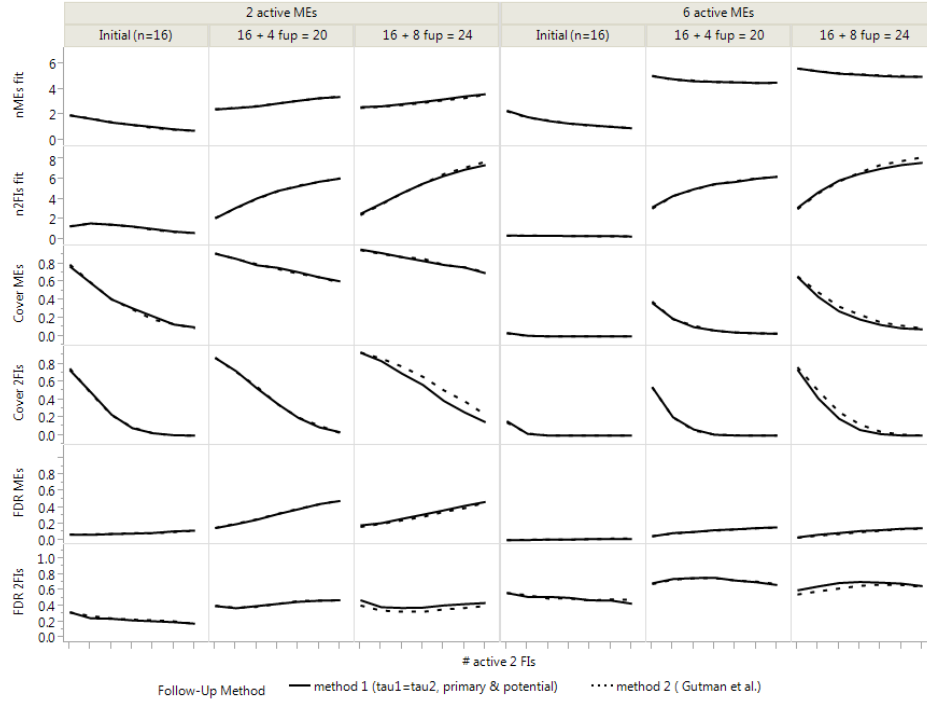
*Figure 5.4 Power and type I error rates for 8-factor case*



### 5.5 Coverages by effect size for 8-factor case

The top two rows of Figure 5.6 illustrate the mean number of main effects and two-factor interaction fit after each set of iterations, while the middle two and final two rows illustrate coverage probabilities and FDR's respectively. The increase in coverage is again apparent when two main actives are active and method 2 is implemented. Further, using either updating method results in the same number of model terms being fit and FDR's either unchanged or slightly reduced, indicating that method 2 increases power and coverage without also increasing FDR's.





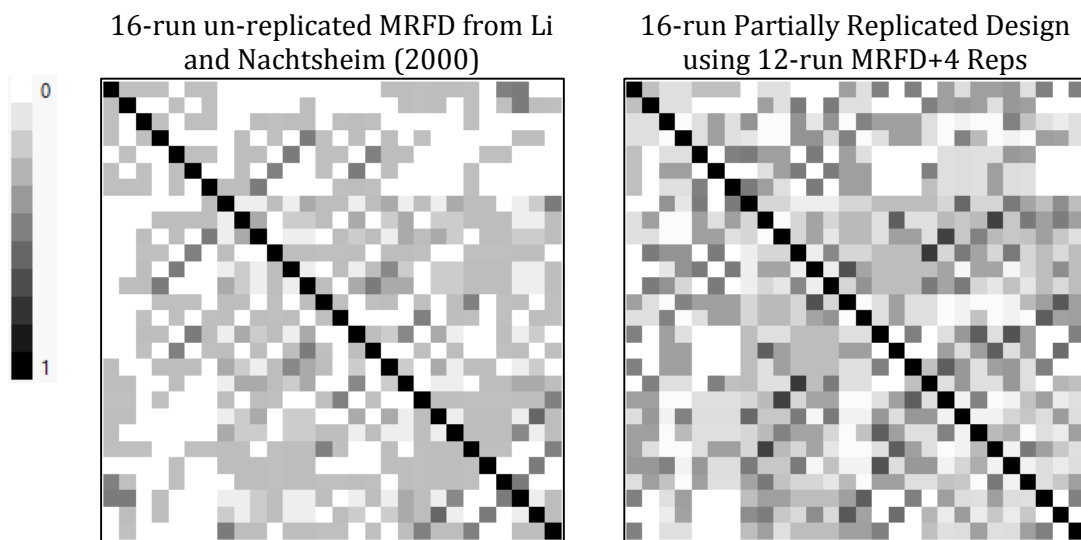
5.6 Number terms fit, cover, and FDR's for 8-factor case

#### 5.4.2 7-factor Designs using $n_i = 16$

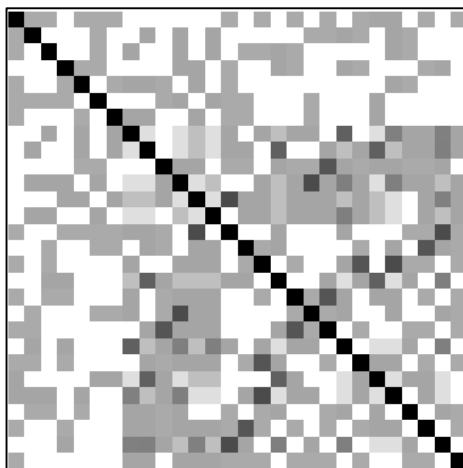
A second simulation study was conducted using 7-factor designs. This was done in order to determine if having more unique runs relative to the number of factors influences whether or not an un-replicated or partially replicated design is initially used. For the second simulation, initial run sizes of  $n_i = 16$  and follow-up run sizes of  $n_f = \{4, 8, 16\}$  are used. Three scenarios were investigated. 1) Beginning with the partially replicated initial design, un-replicated follow-up runs were used in all three settings. 2) Beginning with an un-replicated initial design, a set of Bayesian *DP*-optimal follow-up runs are used for all three follow-up settings. 3) Beginning with an un-replicated initial design, a fixed number of replicate follow-up runs are used when  $n_f = 8$  or 16, where four of the 8 or 16 follow-up runs are forced to be replicate points using the same method as in the first simulation. Only

results when implementing the second updating method are presented since the updating scheme benefits the Bayesian  $DP$ -optimal follow-up procedure without negatively affecting the forced or Bayesian  $D$ -optimal approaches.

The un-replicated 7-factor MRFD's from Li and Nachtsheim (2000) and the 12-run MRFD design augmented with four replicates using the Bayesian  $DP$ -criterion are provided in Table 5.3. Correlation maps of the two 16-run designs are shown in the top two plots in Figure 5.7. The correlation map of the 12 unique runs from the partially replicated design is shown in the lower of the three plots. There are correlations among the main effects in all three designs. For the 16-run un-replicated design, all non-zero correlations among the main effects are of 0.25. For the partially replicated design, correlations among main effects range from zero up to a single correlation of 0.5 between  $x_5$  and  $x_7$ . Compared with the original 12-run design shown the lower plot, a number of correlations among the main effects appear to be increased; however, when considering the estimation of a main-effects only model,  $D_{eff}$  is shown to increase from 0.86 to 0.87 after adding the four replicate points. Also, visual inspection of the plots seems to suggest that correlations between the main effects and two-factor interaction also appear to be increased with the addition of the four replicate points to the 12-run design; however,  $tr(\mathbf{AA}')$  is shown to only slightly increase from 37.52 to 37.61 when adding the replicate points, indicating that the potential impact from model-misspecification is not increased due to the addition of the replicate runs.



12-run MRFD used to create Partially Replicated Design



*Figure 5.7 Correlation maps for 7-factor initial designs*

Table 5.3 Un-replicated and partially replicated 8-factor MRFD's

Un-replicated MRFD								Partially Replicated MRFD							
$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$	$x_8$	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$	$x_8$
-1	1	1	-1	1	1	-1	-1	1	-1	1	-1	-1	1	-1	1
-1	1	-1	1	-1	1	-1	-1	-1	1	1	1	-1	1	1	-1
-1	-1	-1	1	1	-1	-1	-1	1	1	-1	-1	1	-1	1	1
1	1	1	1	1	1	1	1	1	-1	-1	1	-1	-1	1	1
-1	-1	1	-1	1	-1	1	-1	-1	1	1	-1	1	1	1	-1
1	-1	-1	1	1	-1	1	1	1	-1	-1	-1	1	-1	-1	1
1	1	-1	-1	-1	-1	-1	1	-1	1	-1	-1	-1	-1	-1	-1
-1	1	1	1	-1	-1	-1	-1	1	-1	1	1	-1	-1	-1	1
1	1	-1	-1	1	1	1	1	-1	-1	1	-1	1	-1	1	-1
1	-1	-1	-1	-1	-1	1	1	1	1	-1	1	-1	1	-1	1
1	-1	1	-1	1	1	-1	1	-1	-1	-1	1	1	1	1	-1
1	1	-1	1	-1	-1	1	1	-1	1	1	1	1	1	-1	-1
-1	-1	1	-1	-1	1	1	-1	-1	1	-1	-1	-1	-1	-1	-1
-1	-1	1	1	-1	1	-1	-1	1	-1	-1	1	1	1	1	-1
1	1	1	-1	-1	-1	-1	1	1	-1	1	-1	-1	1	-1	1
-1	-1	-1	1	1	1	1	-1	1	1	-1	-1	1	-1	1	1

Number of terms fit, coverages, and FDR's for both main effects and two-factor interactions are illustrated in Figure 5.10. The almost identical results indicate that beginning with an initially un-replicated or partially replicated 16-run 7-factor design does not influence final results when four  $df_{pe}$  are to be used for estimating pure error. Concerning the shared performance, results show that when effect sparsity holds in the main effects (2 of 6 MEs active), follow-up procedures allow for maximum coverage of main effects when the run size is doubled and models are, on average, overfit with 1 of 3 main effects incorrectly identified as active. However, minimized FDR's of dominant effects in Figure 5.9 indicate that the incorrectly identified main effects are large in magnitude. Also, although coverage for two-factor interactions is not maximized even with a doubling of the initial run size as follow-up, power for two-factor interactions is consistently around 0.80.

Figure 5.9 shows maximum coverage and minimum FDR's of dominant effects when following up with 16 runs even when effect sparsity is challenged in the main effects (6 of 7 active). Figure 5.10 shows coverage, FDR's, and power for effects of moderate effect size. These results show using more than eight follow-up runs are needed in order to minimize FDR's for moderate effects and that a 16-run follow-up can maximize power to detect effects of this size although maximizing coverage requires either more runs or a different overall experimental strategy. Analyzing results by effect size also helps to explain the low coverage in two-factor interactions. Since dominant and moderate sized effects are well covered with minimal error, the undetected two-factor interactions are likely small in magnitude.

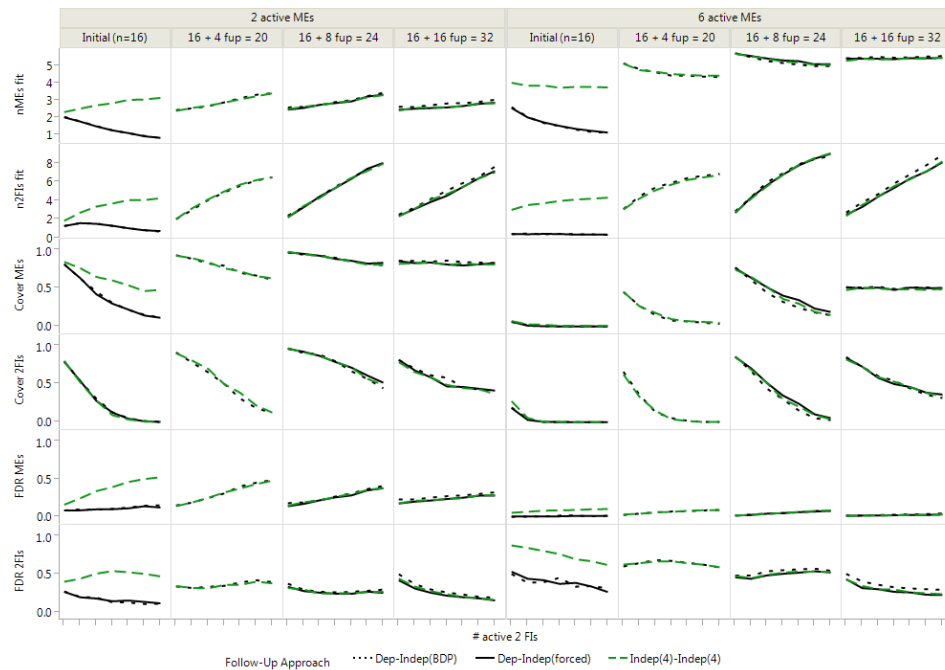


Figure 5.8 Number terms fit, cover, and FDR's for 7-factor case

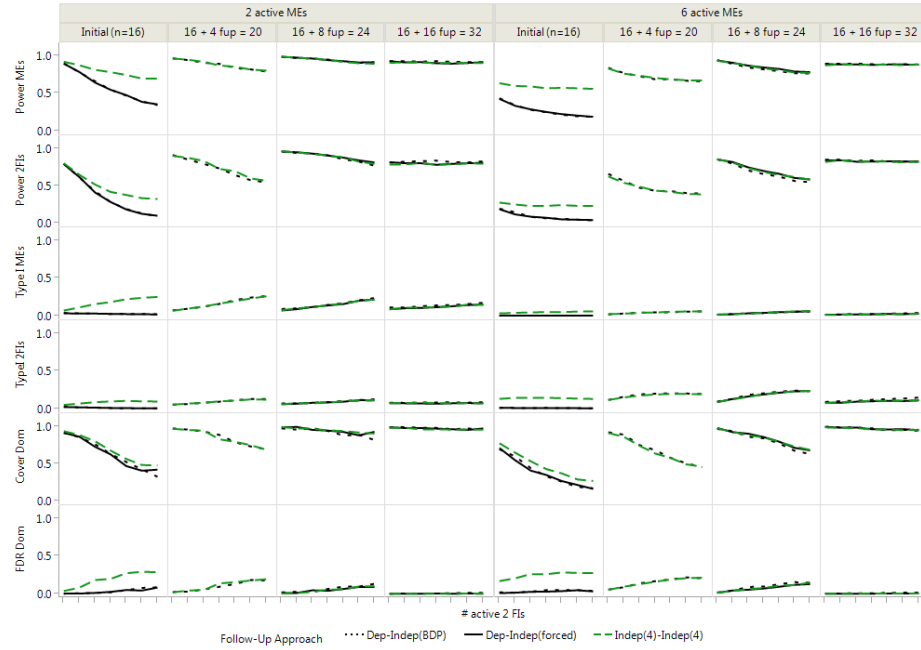


Figure 5.9 Power, type I error, FDR's for 7-factor case

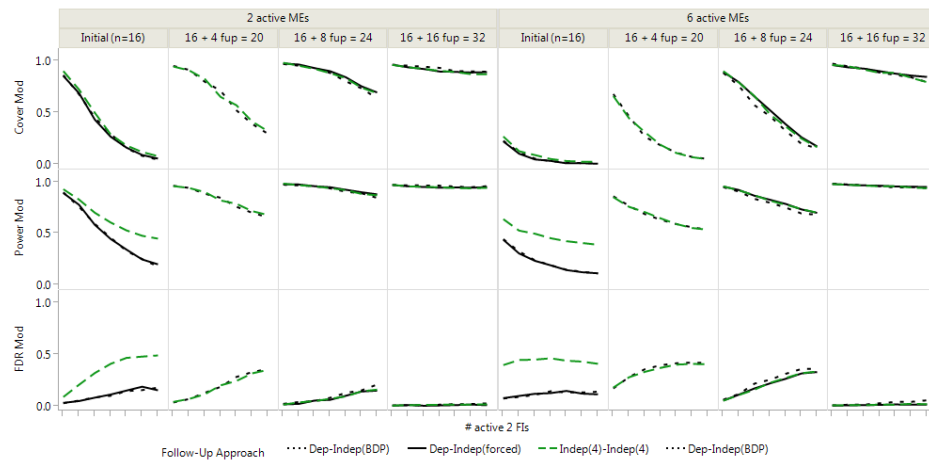


Figure 5.10 Cover, power, and FDR's for moderate effect sizes for 7-factor case

## 5.5 Comments and Discussion

In this chapter we investigated following-up 8-factor and 7-factor screening designs. Un-replicated and partially replicated 16-run MRFD's were used to investigate performance of overall screening strategies when implementing forward selection with a model-independent error estimate. The augmentation approach of Gutman et al. (2014) was adapted to the supersaturated model case where initial screening of main effects and two-factor interactions are considered. Also, the Bayesian *DP*-optimality criterion was extended to augment initial screening designs with partially replicated follow-up runs. Using weak effect heredity and results from the initial experiment to specify secondary terms was found to improve follow-up approaches when implementing Bayesian *DP*-augmentation. This was due, in part, to the updating method influencing the criterion to deem a fewer number of replicate points optimal and therefore permitting more unique design points to be formed. Similar to the *BDP* designs presented in earlier chapters, the complete 7- and 8-factor designs used in this chapter when implementing the Bayesian *DP*-optimality criterion contain between five to seven replicate points.

Once follow-up runs and analyses were carried out, simulation results did not indicate a change in performance when using an un-replicated versus a partially replicated initial screening design. Further research is needed to better understand the defining of prior distributions for follow-up experiments when initial knowledge of main effects and their two-factor interactions is poor. A more structured design approach is planned for this future investigation, where possible factors could include: initial design size, follow-up run size, number of replicates in initial and follow-up designs, complexity of initial model assumption,

number of follow-up experiments, and using different representations for the value of  $p$  in the calculation of the BDP-criterion. Also, different underlying models are to be investigated, such as the null case, supersaturated designs, quadratic models, and changes in heredity assumptions.



## Chapter 6

### Conclusion and Final Comments

Many approaches found in the literature present un-replicated and partially replicated designs based on some criterion for a given number of factors and run size without much or any practical support of the 'high/low', 'maximized/minimized', or targeted criterion value translating to actual, expected performance. This build-up of a number of competing screening designs (both un-replicated & partially replicated) can be attributed to the need for computationally intensive computer simulation to assess the true potential of these designs. However, due to the seemingly endless number of designs, levels of model complexity, and analysis procedure available, simulation studies are limited in their ability to be generalized.

In this dissertation we present simulation studies that can be considered generalizable to the extent of considering first order with two-factor interaction models where the number of potentially active effects fall into the range presented. Allowing the true underlying model to be constructed via random assignment of active effects using all columns of the design provided a more representative simulated screening environment. Examining results based upon effect size allowed competing screening methodologies to be assessed in terms of detecting the more influential effects (dominant, moderate). This gave more insight into overall performance than focusing on the detection of main effects which may or may not be of relative importance if some two-factor interactions are larger in magnitude than some active, though less influential, main effects.

In Chapter 3 simulation results show that the degrading performance of analysis procedures when paired with partially replicated designs may correspond to abrupt changes in  $tr(\mathbf{AA}')$  values. This connection was made by combining information gathered from investigating design criteria with patterns found in the simulation results. The newly proposed Bayesian *DP*-criterion has been shown to coincide with this change point by providing a possible upper bound on  $tr(\mathbf{AA}')$  when adding replicate runs to a design. However, designs created using a two-step construction procedure were found in some cases to be more BDP-efficient than the BDP-'optimal' design created via a candidate-exchange algorithm and using the actual Bayesian *DP*-criterion. Further, although an abrupt change in  $tr(\mathbf{AA}')$  was apparent for some cases, investigation of a more saturated model case showed increases in  $tr(\mathbf{AA}')$  to be almost linear, therefore making the findings less generalizable.

Simulation results from Chapter 3 also illustrate that the choice of screening design should include considerations for replicating some of the design points. This consideration stems from results showing that forward selection using a model-independent error estimate provides comparable performance in some cases to the Dantzig selector, which has been shown in the literature to outperform many other analysis procedures in simulation. Forward selection was shown to benefit from the incorporation of a model-independent error estimate, with increases in power to detect active effects, an improved ability to form more properly sized models (i.e., correct number of main effects and two-factor interactions), and an increased ability to detect effects of all sizes. This was at the expense of higher FDR's and type I errors being committed, although in most cases FDR's of dominant effects were still kept to minimal levels.

The Dantzig selector was found to generally perform better when combined with an un-replicated design. The main advantage appeared to be in the ability of the procedure to properly detect two-factor interactions; however, there were two designs found in Chapters 3 and 4 which would not permit the Dantzig selector to fit any two-factor interactions. Further, combining one of these two designs with forward selection not only showed higher power to detect main effects with minimal type I error when compared to the Dantzig selector, forward selection was also shown to provide acceptable power in detecting two-factor interactions with minimal type I errors rates, therefore illustrating the importance of investigating the relationships between design of experiments and model selection procedures.

In Chapter 4, the Bayesian *DP*-optimality criterion was incorporated into the Pareto front approach to investigate the possibility of further reducing the potential impact from model-misspecification. PAPE constructed designs were not shown to outperform designs constructed from simply using a single-criterion approach; however, the multi-criterion approach does permit for a wider selection of designs to be further compared based upon additional criteria. Although in the work presented here only  $df_{pe}$  and EC were considered, the hundreds of competing designs on the Pareto front could be compared and reduced to a smaller set using other criteria, such as model discrimination type metrics, which provide a more direct connection to analysis procedure.

In Chapter 5 the Bayesian *DP*-optimality criterion was incorporated into follow-up experimentation, which allowed for the determination of an optimal set of partially replicated follow-up runs. Simulation studies focusing on the forward selection procedure showed that there was no difference in screening performance whether one chose to begin

with a partially replicated design or if one began with an un-replicated design and followed-up with a partially replicated set of runs. The incorporation of weak effect heredity into the updating of prior information by including particular two-factor interactions as secondary terms, showed some indication of improving overall sequential screening performance. The improvement was small and was found to be related to the Bayesian  $DP$ -optimality criterion adjusting the amount of replication deemed optimal when few terms were deemed active.

More work is planned for simulating sequential experimentation. In particular, a more structured approach will be used by implementing designed experiments. This will require the determination of factors and their levels in addition to forming a response metric more practical for analysis purposes. Using this approach, further investigation will be made into the abrupt changes found in  $tr(\mathbf{A}\mathbf{A}')$  as the number of replicates contained in a design increases. Because these changes in  $tr(\mathbf{A}\mathbf{A}')$  can be identified without simulation, a possible design space can be formed and simulation used to verify the relationships found, thereby saving both time and resources. Results from this dissertation suggest that the Bayesian  $DP$ -optimality criterion can in some cases closely correspond to this change. Future work will include improving upon this criterion, and investigating the properties and performance of designs created using the criterion.

## Reference List

- Abraham, B., Chipman, H. and Vijayan, K. (1999). Some Risks in the Construction and Analysis of Supersaturated Designs. *Technometrics*, 41, 135-141.
- Allen, T.T., Yu, L. and Schmitz, J. (2003). An Experimental Design Criterion for Minimizing Meta-Model Prediction Errors Applied to Die Casting Process Design. *Journal of the Royal Statistical Society, Ser. C*, 52, 103-117.
- Anderson-Cook, C.M. and Hamada, M.S. (2014). Comment: Toward Guidelines for Practitioners on Screening Designs and Analysis. *Technometrics*, 56, 16-19.
- Atkinson, A.C., Donev, A.N. and Tobias, R.D. (2007). *Optimum Experimental Designs, with SAS*. (Vol. 34). Oxford University Press, Oxford.
- Booth, K.H. and D.R. Cox (1962). Some Systematic Supersaturated Designs. *Technometrics*, 4, 489-495.
- Brenneman, W.A. (2014). Comment: Simulation Used to Solve Tough Practical Problems. *Technometrics*, 56, 19-20.
- Bursztyn, D. and Steinberg, D.M. (2006). Comparison of Designs for Computer Experiments. *Journal of Statistical Planning and Inference*, 136, 1103-1119.
- Candes, E.J. and Tao, T. (2007). The Dantzig Selector: statistical estimation when  $p$  is much larger than  $n$ . *Annals of Statistics*, 35, 2365-2369.
- Dasgupta, N., Jacroux, M. and ShaRay, R. (2010). Partially Replicated Fractional Factorial Designs. *Metrika*, 71, 295-311.
- Draguljic, D., Woods, D.C., Dean, A.M., Lewis, S.M. and Vine, A.J.E. (2014). Screening Strategies in the Presence of Interactions. *Technometrics*, 56, 1-16.
- Draper, N.R. and Smith, H. (1998). *Applied Regression Analysis*, 3rd Ed. John Wiley and Sons, New York, NY.
- DuMouchel, W. and Jones, B. (1994). A Simple Bayesian Modification of  $D$ -Optimal Designs to Reduce Dependence on an Assumed Model. *Technometrics*, 36, 37-47.

- Edwards, D.J. and Kelly, K. (2014). Selecting a  $D$ -Optimal Follow-Up Experiment among Candidate Choices. *Quality Engineering*, 26, 233-245.
- Edwards, D.J. and Mee, R.W. (2011). Supersaturated Designs: Are Our Results Significant? *Computational Statistics and Data Analysis*, 55, 2652-2664.
- Edwards, D.J., Weese, M.L. and Palmer, G.A. (2014). Comparing Methods for Design Follow-Up: Revisiting a Metal-Cutting Case Study. *Applied Stochastic Models in Business and Industry*, 30, 464-478.
- Fedorov, V.V. (1972). *Theory of Optimal Experiments*. Academic Press, New York, NY.
- Gilmour, S. and Trinca, L.A. (2012). Optimum design of experiments for statistical inference. *Journal of the Royal Statistical Society, Applied Statistics, Part 3*, 61, 345-401.
- Goos, P. and Jones, B.A. (2011). *Optimal Design of Experiments: A Case Study Approach*. Wiley and Sons, New York, NY.
- Gutman, A.J., White, E.D., Lin, D.K.J. and Hill, R.R. (2014). Augmenting Supersaturated Designs with Bayesian  $D$ -Optimality. *Computational Statistics and Data Analysis*, 71, 1147-1158.
- Jones, B. and Nachtsheim, C.J. (2011). A Class of Three-Level Designs for Definitive Screening in the Presence of Second-Order Effects. *Journal of Quality Technology*, 43, 1-15.
- Jones, B., Lin, D.K.J. and Nachtsheim, C.J. (2008). Bayesian  $D$ -Optimal Supersaturated Designs. *Journal of Statistical Planning and Inference*, 138, 86-92.
- Lin, D.K.J. (1993). A New Class of Supersaturated Designs. *Technometrics*, 35, 28-31.
- Li, W. Lin, D.K.J. (2003). Analysis Methods for Supersaturated Design: Some Comparisons. *Journal of Data Science*, 1, 249-260.
- Li, W. and Nachtsheim, C.J. (2000). Model-Robust Factorial Designs. *Technometrics*, 42, 345-352.
- Liao, C.T. and Chai, F.S. (2009). Design and Analysis of Two-Level Factorial Experiments with Partial Replication. *Technometrics*, 51, 66-74.
- Lu, L., Anderson-Cook, C.M. and Robinson, T.J. (2011). Optimization of Designed Experiments Based on Multiple Criteria Utilizing a Pareto Frontier. *Technometrics*, 53, 353-365.
- Lu, L. and Anderson-Cook, C.M. (2012). Rethinking the Optimal Response Surface Design for a First-Order Model with Two-Factor Interactions, When Protecting against Curvature. *Quality Engineering*, 24, 404-422.

- Lupinacci, P.J. and Pigeon, J.G. (2008). A Class of Replicated Two Level Fractional Factorial Designs. *Journal of Quality Technology*, 40, 184-193.
- Marley, C.J. and Woods, D.C. (2010). A Comparison of Design and Model Selection Methods for Supersaturated Experiments. *Computational Statistics and Data Analysis*, 54, 3158-3167.
- Mee, R.W. (2013). Tips for Analysing Nonregular Fractional Factorial Experiments. *Journal of Quality Technology*, 45, 330-349.
- Mee, R.W., Schoen, E.D. and Edwards, D.J. (2015). Selecting Nonregular Factorial Designs for Screening, submitted to *Technometrics*.
- Meyer, R.K. and Nachtsheim, C.J. (1995). The Coordinate Exchange Algorithm for Constructing Exact Optimal Experimental Screening Designs. *Technometrics*, 37, 60-69.
- Mitchell, T.J. (1974). An Algorithm for the Construction of D-Optimal Experimental Designs. *Technometrics*, 16, 203-210.
- Montgomery, D.C. (2009). *Design and Analysis of Experiments*, 7th Edition. John Wiley and Sons, Hoboken, NJ.
- Myers, R.H. (1990). *Classical and Modern regression with Applications*, 2nd Edition. Wadsworth Publishing, Belmont, CA.
- Myers, R.H., Montgomery, D.C. and Anderson-Cook, C.M. (2009), *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, 3rd Edition. New York, NY, John Wiley and Sons.
- Neff, A. (1996). Bayesian Two-Stage Designs Under Model Uncertainty. Ph.D. Thesis, Virginia Polytechnic and State University, Virginia.
- Ou, Z., Qin, H. and Cai, X. (2013). Partially Replicated Two-Level Fractional Factorial Designs via Semifoldover. *Journal of Statistical Planning and Inference*, 143, 809-817.
- Phoa, F.K.H, Pan, Y. and Xu, H. (2009). Analysis of supersaturated designs via the Dantzig Selector. *Journal of Statistical Planning and Inference*, 139, 2362-2372.
- Robinson, T.J. and Anderson-Cook, C.M. (2011). A Closer Look at *D*-Optimality for Screening Designs. *Quality Engineering*, 23, 1-14.
- Ruggoo, A. and Vandebroek, M. (2004). Bayesian Sequential *D-D* Optimal Model-Robust Designs. *Computational Statistics and Data Analysis*, 47, 655-673.
- Scinto, P.R., Wilkinson, R.G., Wang, Z. and Rose, A. (2014). Comment: Need for Guidelines on Appropriate Screening Designs for Practitioners. *Technometrics*, 56, 23-24.

Silvestrini, R.T. (2013). Considerations for *D*-Optimal Sequential Design. *Quality and Reliability Engineering International*, 31, 399-410.

Smith, K. (1918). On the Standard Deviations of Adjusted and Interpolated Values of an Observed Polynomial Function and Its Constants and the Guidance They Give Towards a Proper Choice of the Distribution of Observations. *Biometrika*, 12, 1-85.

Sunburg, R. (2008). A Classical Dataset from Williams, and Its Role in the Study of Supersaturated Designs. *Journal of Chemometrics*, 22, 436-440.

Schwarz, G. (1978). Estimating the Dimension of a Model. *Annals of Statistics*, 6, 461-464.

Tsai, S.F. and Liao, C.T. (2014). Selection of Partial Replication on Two-Level Orthogonal Arrays. *The Canadian Journal of Statistics*, 42, 168-183.

Wald, A. (1943). On the Efficient Design of Statistical Investigations. *Annals of Mathematical Statistics*, 14, 134-140.

Wang, P.C. (1993). Comments on Lin. *Technometrics*, 37, 358-359.

Westfall, P.H., Young, S.S. and Lin, D.K.J. (1998). Forward Selection Error Control in the Analysis of Supersaturated Designs. *Statistica Sinica*, 8, 101-117.

Wu, J.C.F. and Hamada, M.S. (2009). *Experiments: Planning, Analysis, and Optimization*, 2nd Edition. Wiley, New York.



# Appendix

## Designs

### Chapter 3

#### 7-factor 20-run designs

D-optimal (un-replicated)	Bayesian D-optimal (un-replicated)	DP-optimal with 12 dfpe
-1 1 1 -1 -1 1 -1	-1 -1 1 -1 -1 -1 1	1 -1 1 -1 -1 1 1
1 -1 1 1 1 -1 -1	1 -1 -1 1 1 -1 1	-1 1 -1 -1 -1 -1 1
-1 -1 1 1 -1 1 -1	1 1 -1 1 1 1 -1	1 -1 -1 1 -1 -1 -1
1 -1 1 1 -1 1 1	-1 1 1 -1 1 1 -1	1 -1 1 -1 -1 1 1
-1 -1 1 -1 1 -1 -1	-1 1 -1 1 1 -1 1	1 1 1 -1 1 -1 -1
1 -1 -1 1 -1 -1 1	-1 -1 -1 -1 -1 -1 -1	1 1 1 -1 1 -1 -1
-1 1 -1 1 -1 1 -1	-1 1 -1 -1 -1 1 1	1 1 -1 1 1 1 1
1 1 1 -1 1 1 1	-1 -1 1 1 1 -1 -1	1 1 -1 1 1 1 1
-1 1 -1 1 1 1 1	-1 -1 -1 1 -1 1 -1	-1 -1 -1 -1 1 1 -1
-1 1 -1 -1 -1 -1 1	1 -1 -1 -1 -1 1 1	1 1 1 -1 1 -1 -1
1 1 1 1 -1 -1 -1	1 1 1 -1 -1 1 -1	-1 1 1 1 -1 1 -1
-1 -1 1 -1 1 1 1	-1 -1 -1 -1 1 1 1	-1 1 1 1 -1 1 -1
-1 -1 -1 1 1 -1 1	1 -1 1 -1 1 1 -1	1 -1 -1 1 -1 -1 -1
1 -1 -1 -1 1 1 -1	1 1 1 -1 1 1 1	1 -1 1 -1 -1 1 1
-1 1 1 1 1 -1 1	1 1 -1 -1 1 -1 -1	-1 -1 -1 -1 1 1 -1
1 1 -1 1 1 1 -1	1 -1 1 1 -1 -1 -1	-1 1 -1 -1 -1 -1 1
1 -1 -1 -1 -1 1 1	-1 -1 1 1 -1 1 1	1 -1 -1 1 -1 -1 -1
1 1 1 -1 -1 -1 1	1 1 1 1 1 -1 1	-1 -1 1 1 1 -1 1
1 1 -1 -1 1 -1 -1	1 1 -1 1 -1 -1 1	-1 1 -1 -1 -1 -1 1
-1 -1 -1 -1 -1 -1 -1	-1 1 1 1 -1 -1 -1	-1 -1 1 1 1 -1 1

Bayesian DP with 6 dfpe	BDBD with 4 dfpe	BDBD with 2 dfpe
1 1 -1 1 -1 -1 -1	-1 -1 -1 1 1 1 1	-1 1 1 -1 -1 1 1
1 1 1 -1 -1 1 1	-1 -1 1 -1 -1 -1 1	1 1 1 -1 1 -1 -1
-1 1 -1 -1 -1 1 -1	1 -1 -1 -1 -1 1 1	-1 1 1 1 -1 -1 -1
-1 -1 -1 1 1 -1 -1	-1 1 -1 -1 -1 1 -1	-1 1 -1 -1 1 -1 -1
1 1 1 1 1 1 -1	1 -1 -1 -1 1 1 -1	1 1 -1 -1 -1 1 -1
-1 -1 1 1 -1 -1 -1	1 1 1 -1 1 1 1	-1 -1 -1 -1 -1 -1 -1
1 -1 -1 -1 -1 1 -1	-1 -1 -1 1 -1 -1 -1	-1 -1 1 -1 1 1 -1
-1 1 1 1 1 -1 1	1 1 1 -1 -1 -1 -1	1 -1 -1 -1 1 1 1
1 -1 1 1 1 -1 1	1 -1 1 1 1 -1 1	-1 -1 1 1 1 -1 -1
1 1 1 -1 1 -1 -1	1 1 -1 1 -1 -1 1	1 1 -1 -1 -1 -1 1
-1 1 1 1 1 -1 1	-1 1 -1 -1 1 -1 1	-1 -1 1 -1 1 -1 1
-1 -1 1 1 -1 -1 -1	1 -1 1 1 -1 1 -1	-1 1 -1 1 1 1 -1
1 1 1 -1 1 -1 -1	1 1 -1 1 1 -1 -1	1 -1 -1 1 1 -1 -1
1 1 -1 1 -1 -1 -1	-1 1 1 1 1 1 -1	1 1 1 1 1 1 1
-1 -1 1 -1 1 1 1	-1 -1 1 -1 1 -1 -1	-1 -1 -1 1 -1 1 1
-1 -1 1 -1 1 1 1	-1 1 1 1 -1 1 1	1 -1 1 1 -1 1 -1
-1 -1 -1 -1 -1 -1 1	1 1 1 -1 -1 -1 -1	1 -1 1 1 -1 -1 1
1 1 1 1 1 1 -1	-1 -1 -1 1 -1 -1 -1	-1 1 -1 1 1 -1 1
-1 -1 -1 1 -1 1 1	1 -1 1 1 1 -1 1	1 1 1 1 1 1 1
1 1 -1 -1 1 1 1	-1 1 -1 -1 1 -1 1	-1 -1 -1 1 -1 1 1

## Chapter 4

### 8-factor 16-run designs

d16 PAPE BD design (un-replicated)	d18 PAPE BD design (replicated)
1 1 -1 -1 -1 -1 1 1	-1 -1 1 -1 -1 -1 1 -1
1 -1 -1 1 -1 1 1 -1	1 1 -1 -1 -1 -1 -1 1
-1 -1 -1 -1 -1 -1 -1 -1	1 1 -1 1 1 -1 1 -1
-1 1 -1 1 -1 1 -1 1	-1 -1 -1 1 -1 1 -1 -1
1 -1 -1 -1 1 1 -1 1	1 -1 1 1 -1 1 1 1
1 1 1 1 1 1 1 1	-1 1 -1 1 -1 1 -1 -1
1 1 -1 1 1 -1 -1 -1	1 -1 1 -1 1 1 -1 -1
1 -1 1 1 -1 -1 -1 1	1 1 1 1 -1 1 1 1
-1 -1 -1 1 1 -1 1 1	-1 -1 1 1 1 -1 -1 1
-1 1 1 -1 1 -1 -1 1	-1 1 1 1 1 -1 -1 1
1 -1 1 -1 1 -1 1 -1	-1 1 1 1 1 1 1 -1
-1 -1 1 1 1 1 -1 -1	1 -1 -1 1 1 -1 1 -1
-1 1 1 1 -1 -1 1 -1	-1 1 -1 -1 1 1 1 1
1 1 1 -1 -1 1 -1 -1	1 1 1 -1 1 1 -1 -1
-1 1 -1 -1 1 1 1 -1	-1 1 1 -1 -1 -1 1 -1
-1 -1 1 -1 -1 1 1 1	-1 -1 -1 -1 1 1 1 1



## Chapter 5

### 8-factor and 7-factor un-replicated and partially replicated MRFD

8-factor Designs in 16 runs															
MRFD from Li and Nachtsheim (2000) un-replicated								12-MRFD with 4 augmented replicate points							
-1	1	-1	1	1	1	1	-1	1	1	-1	-1	1	1	-1	1
1	1	1	1	1	-1	-1	-1	1	-1	-1	1	-1	1	1	-1
1	1	1	1	-1	1	1	1	1	1	1	1	-1	-1	-1	-1
1	-1	1	-1	-1	-1	-1	-1	-1	1	1	-1	1	1	-1	-1
-1	-1	-1	-1	1	1	1	1	-1	-1	1	1	-1	-1	1	-1
1	-1	-1	1	1	1	-1	1	1	1	-1	1	-1	-1	1	1
-1	1	-1	1	-1	1	1	-1	-1	1	1	1	1	1	1	-1
-1	-1	-1	-1	1	-1	-1	-1	-1	-1	-1	1	-1	1	1	1
-1	-1	1	1	-1	-1	-1	1	-1	-1	-1	-1	-1	-1	-1	1
1	-1	1	1	-1	1	1	-1	1	1	-1	-1	-1	-1	-1	1
-1	1	-1	-1	-1	-1	-1	1	1	-1	-1	-1	-1	-1	-1	1
1	1	-1	-1	1	1	1	1	-1	-1	-1	-1	1	1	1	-1
1	1	1	-1	1	-1	1	-1	1	1	-1	-1	1	1	1	1
-1	-1	1	1	-1	-1	1	-1	-1	1	1	1	1	1	-1	-1
1	-1	-1	-1	-1	1	-1	1	1	-1	1	1	-1	-1	-1	1
-1	1	1	-1	1	1	-1	1	-1	-1	-1	-1	1	-1	1	-1

7-factor Designs in 16 runs															
MRFD from Li and Nachtsheim (2000) un-replicated								12-MRFD with 4 augmented replicate points							
-1	1	1	-1	1	1	1	-1	1	-1	1	-1	-1	1	1	-1
-1	1	-1	1	-1	1	1	-1	-1	1	1	1	-1	1	1	1
-1	-1	-1	1	1	1	-1	-1	1	1	-1	-1	1	-1	1	1
1	1	1	1	1	1	1	1	1	-1	-1	1	-1	-1	1	1
-1	-1	1	-1	1	1	-1	1	-1	1	1	1	1	1	1	1
1	-1	-1	1	1	1	-1	1	1	-1	-1	-1	1	-1	-1	-1
1	1	-1	-1	-1	-1	-1	-1	-1	1	-1	-1	-1	-1	-1	-1
-1	1	1	1	-1	-1	-1	-1	1	-1	1	1	-1	-1	-1	-1
1	1	-1	-1	1	1	1	1	-1	-1	1	-1	1	-1	1	1
1	-1	-1	-1	-1	-1	-1	1	1	1	-1	1	1	1	1	1
1	1	-1	1	-1	-1	-1	1	-1	1	1	1	1	1	-1	-1
-1	-1	1	-1	-1	1	1	1	-1	1	-1	-1	-1	-1	-1	-1
-1	-1	1	1	-1	1	1	-1	-1	-1	-1	1	1	1	1	1
1	1	1	-1	-1	-1	-1	-1	1	-1	1	-1	-1	1	-1	-1
-1	-1	-1	1	1	1	1	1	1	1	-1	-1	1	-1	1	1

## Vita

Robert D. Leonard was born on May 25, 1980, in Muskegon, Michigan, and is an American citizen. He graduated from Oak Hill High School, in Oak Hill, Ohio in 1998. He received his Bachelor of Science in Integrated Mathematics Education from Ohio University in Athens, Ohio in 2003. He taught English as a Second Language in Peru for two years and subsequently taught in public schools in Virginia for three years. He also served as a doctoral research assistant at the NASA Langley Research Center while working on his Ph.D. at Virginia Commonwealth University.